EFFICIENT COMPUTATION IN STOCHASTIC OPTIMIZATION AND SIMULATION: DYNAMIC PROGRAMS, RISK QUANTIFICATION, AND RISK OPTIMIZATION

A Thesis
Presented to
The Academic Faculty

by

Helin Zhu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2016

EFFICIENT COMPUTATION IN STOCHASTIC OPTIMIZATION AND SIMULATION: DYNAMIC PROGRAMS, RISK QUANTIFICATION, AND RISK OPTIMIZATION

Approved by:

Professor Enlu Zhou,
Committee Chair
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Hayriye Ayhan
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Enlu Zhou, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Shijie Deng
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Sigrún Andradóttir
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Henry Lam
Department of Industrial and Operations Engineering
University of Michigan

Date Approved: 21 July 2016

This thesis is dedicated to my parents, my sister and my wife, for their unconditional love and support throughout my PhD study.

I would like to express my deepest gratitude to my advisor, Dr. Enlu Zhou, for her continuous guidance, supervision, and support on the projects, and mentorship on my life. I could not imagine having a better advisor.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Professor Enlu Zhou, Professor Sigrún Andradóttir, Professor Hayriye Ayhan, Professor Shijie Deng, and Professor Henry Lam for serving as my PhD thesis committee members. I am very grateful for their insightful questions, comments and suggestions that help improve my research and thesis in various aspects.

I would also like to thank my fellow group members, Dr. Fan Ye, Joshua Hale, Dr. Chang-Han Rhee, Di Wu, as well as many other professors and PhD students in ISyE for the countless discussions during my research, and their valuable contributions to my research and thesis.

TABLE OF CONTENTS

DEDICATION . . . . . **iii**

ACKNOWLEDGEMENTS . . . . . **iv**

LIST OF TABLES . . . . . **viii**

LIST OF FIGURES . . . . . **ix**

SUMMARY . . . . . **x**

I INTRODUCTION **1**

1.1 Solving the Dual Problems of Dynamic Programs via Regression . . 1

1.1.1 Literature Review 3

1.1.2 Motivation and Research Goals 6

1.1.3 Research Results and Contributions 7

1.2 Assessing Risk in Stochastic Simulation under Input Uncertainty . . 9

1.2.1 Motivation and Research Goals 11

1.2.2 Research Results and Contributions 13

1.3 Optimizing Risk Measures via Gradient-based Stochastic Adaptive Search 15

1.3.1 Motivation and Research Results 17

1.4 Outlines 18

II SOLVING THE DUALS OF DYNAMIC PROGRAMS VIA REGRESSION **19**

2.1 Dynamic Programs and Dual Formulations 19

2.1.1 A Dual Formulation 21

2.2 A Regression Approach 25

2.2.1 General Framework 25

2.2.2 Regression Approach with L^2 Orthonormal Basis 27

2.2.3 Regression Approach with Taylor Series Basis 34

2.2.4 Implementation 37

2.3 Special Cases . . . 39
2.3.1 American Option Pricing . . . 39
2.3.2 Controlled Markov Diffusion . . . 41
2.3.3 Linear Quadratic Control . . . 42
2.4 Numerical Experiments: Dynamic Trading . . . 43
2.4.1 Dynamic Trading with Predictable Returns and Transaction Costs . . . 43
2.4.2 Dual Formulation with Regression-based Penalties . . . 47
2.4.3 Numerical Results . . . 49
2.5 Conclusion . . . 53

III RISK ASSESSMENT IN STOCHASTIC SIMULATION UNDER INPUT UNCERTAINTY . . . 55

3.1 Risk Measures of Mean Response . . . 55
3.1.1 A Risk Formulation . . . 55
3.1.2 Nested Monte Carlo Estimators . . . 57
3.2 Asymptotical Analysis . . . 59
3.2.1 Consistency . . . 59
3.2.2 Asymptotic Normality . . . 63
3.3 Budget Allocation . . . 68
3.4 Numerical Experiments . . . 73
3.5 Conclusion . . . 80

IV OPTIMIZATION OF RISK MEASURES VIA GRADIENT-BASED STOCHASTIC ADAPTIVE SEARCH . . . 81

4.1 General Framework . . . 81
4.1.1 Main Idea . . . 82
4.1.2 Review of GASS . . . 84
4.1.3 Extension of GASS to Optimization of CVaR . . . 87
4.2 Algorithms: GASS-CVAR and GASS-CVAR-ARL . . . 88
4.2.1 GASS with Adaptive Risk Level . . . 90

4.3 Convergence Analysis 94
4.4 Numerical Experiments 100
4.5 Conclusion and Future Research 105
APPENDIX A — ADDITIONAL RESULTS IN CHAPTER II . . 107
APPENDIX B — PROOFS OF THEOREMS IN CHAPTER III . 110
APPENDIX C — PROOFS OF THEOREMS IN CHAPTER IV . 125
REFERENCES 131

LIST OF TABLES

2.4.1 Dual Bounds of PLQC Policy. 52

3.4.1 Comparison of Two 95% CI Procedures—VaR. 74

3.4.2 VaR (with 95% CI) for the Mean Sojourn Time in an M/M/1 Queue. 76

3.4.3 CVaR (with 95% CI) for the Mean Sojourn Time in an M/M/1 Queue. 77

A.0.1Dual Bounds of PLQC Policy with Different Φ. 108

A.0.2Dual Bounds of PLQC Policy with Different λ. 109

LIST OF FIGURES

3.4.1 VaR/CVaR CI Width: Pilot Run V.S. Actual Run. 78

4.4.1 Average Performance of GASS-CVaR and GASS-CVaR-ARL. 104

4.4.2 CVaR at Mean of the Sampling Distribution and Trajectory of Risk Level. 105

SUMMARY

Stochastic optimization and simulation are two of the most fundamental research areas in Operations Research. In this thesis, we develop efficient computational approaches for three important topics in the realm of stochastic optimization and simulation.

First, for general dynamic programs, we propose a regression approach to solve the information relaxation dual problems by exploring the structure of the function space of dual penalties. Compared with most of the existing approaches, the proposed one is more efficient since it circumvents the issue of nested simulation in approximating the so-called optimal dual penalty. The resulted approximations maintain to be feasible dual penalties, and thus yield valid dual bounds on the optimal value function. We further apply the proposed framework to a high-dimensional dynamic trading problem to demonstrate its effectiveness and efficiency in solving the duals of complex dynamic programs.

Second, for general stochastic simulation, we study the risk quantification of mean response under input uncertainty, which, to the best of our knowledge, has been rarely systematically studied in the literature. We develop nested Monte Carlo estimators for risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) of mean response under input certainty. We show that they are strongly consistent and asymptotic normally distributed, and thus yield asymptotically valid confidence intervals. We further study the associated budget allocation problem for efficient simulation of the proposed nested risk estimators.

Last, for general loss distributions, we study the extension of the recently proposed model-based approach, namely the gradient-based adaptive stochastic search

(GASS), to the optimization of risk measures such as VaR and CVaR. This problem is usually difficult, because 1) the loss function might lack structural properties such as convexity or differentiability since it is often generated via black-box simulation of a stochastic system; 2) evaluation of VaR or CVaR for a general loss distribution often requires rare-event simulation, which is computationally expensive. Instead of optimizing VaR or CVaR at target risk level directly, we incorporate an adaptive adjustment scheme on the risk level, by initializing the algorithm at a small risk level and adaptively increasing it until the simultaneous achievement of target risk level and convergence of the algorithm. This enables us to adaptively reduce the number of samples needed to estimate the risk measures at each iteration, and thus improving the overall efficiency.

CHAPTER I

INTRODUCTION

Stochastic optimization and simulation are two of the most fundamental and active research areas in Operations Research with extensive applications in modeling and solving real-world problems under uncertainties. In this thesis, we will develop new approaches to study three important topics in the realm of stochastic optimization and simulation. In particular, first we propose a regression approach to solving the information relaxation dual problems of general dynamic programs in stochastic optimization, in which we demonstrate that the optimal dual penalties of general dynamic programs could be approximated efficiently and accurately. Second, we propose a nested simulation approach to assess the risk of simulation output under input uncertainty, in which we show that the proposed nested risk estimators exhibit nice statistical properties such as consistency and asymptotic normality. Lastly, we propose a model-based approach to optimizing risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) for general loss distributions, where we demonstrate that incorporating an adaptive adjustment scheme on the risk level could be extremely beneficial for improving efficiency.

1.1 Solving the Dual Problems of Dynamic Programs via Regression

Many real-world problems could be viewed as sequential decision-making problems in the presence of uncertainties, where the decisions could bring rewards or incur immediate costs and might also impact on the future states of the problems by influencing the probabilistic dynamics. Usually the goal is to find an optimal decision

policy/strategy that maximizes the accumulated rewards (or minimizes the accumulated costs, we will consider maximization of rewards in the rest of the thesis) in the long run among all exercisable decision polices, where a policy provides the decision maker (DM) the actions to take under all possible states of the problems.

Under a discrete-time setting, Markov decision process (MDP) is one of the powerful models for complex sequential decision making problem under uncertainties. For the last several decades, it has been studied extensively theoretically (see, e.g., [66], [14] and [65]), as well as having been applied to problems in various fields such as supply chain management [76], neural networks [13], financial engineering [10], and health care [1]. Optimal policies of general MDPs could be theoretically solved via the classic Bellman backward dynamic programming approach; however, in practice, the dynamic programs (DP) usually suffer from the so-called "curse of dimensionality" [11], meaning that the size of the state space and the action space, and hence the complexity of the program increases exponentially in the dimension of the problem. Therefore, it is rarely the case that the optimal policy of a real-world dynamic programming problem could be solved exactly. Facing this issue, there is abundant existing literature that focuses on developing good approximate dynamic programming methods that aim to construct good suboptimal policies, see [33], [14], [26], [65], etc. In principle, given a suboptimal policy, Monte Carlo simulation could be used to evaluate the policy and generate good lower bound estimators on the optimal value function by simulating a large number of state-action sample paths under the policy. It scales well with the dimension of the underlying system. However, in lack of the exact optimal value function or its upper bounds, the quality of the policy and the optimality gap of the lower bounds are difficult to measure.

The duality theory developed independently by [69] and [24] addresses this issue by formulating and solving the dual representation of the (primal) DP, which generate upper bounds on the optimal value function. If the duality gap, i,e., the difference

between the lower bound induced by the policy and the upper bound, is small enough, then one could claim that the policy is sufficiently good. The main idea of this duality theory is to relax the non-anticipativity constraint on the feasible policies of the DP, i.e., allow the DM to choose actions based on the outcomes of future uncertainties, and penalize the DM for the access to the future information. Thus, this framework is also termed as information relaxation duality theory. In practical implementation, generating an upper bound on the optimal value function using the dual formulation only requires solving multiple scenario-based pathwise deterministic inner optimization problems, and scales well with the dimension of the system via Monte Carlo simulation.

Similar to the duality theory in deterministic (convex) optimization, weak duality and strong duality also hold for the information relaxation duality theory in dynamic programming. In particular, solving the dual DP with a feasible dual penalty generates an upper bound on the optimal value function (weak duality); moreover, if the so-called "optimal dual penalty" is plugged in the dual problem, then the resulted upper bound is tight, meaning that the optimal value function is recovered and strong duality holds. Ideally, to minimize the duality gap, one prefers to compute the optimal dual penalty exactly; however, the optimal dual penalty involves optimal value functions and hence are not readily available. Thus, approximation of the optimal dual penalty is usually required.

1.1.1 Literature Review

The information relaxation duality theory for general DPs originates from the dual theory in pricing American-style options, developed independently by [70], [44], and [3]. They are able to generate upper bounds on the option price by solving the associated dual problem, which is obtained by relaxing the non-anticipativity constraint on all the feasible exercising strategies (which essentially are stopping times) and

penalizing the payoff function with a martingale adapted to the nature information filtration. Furthermore, if the penalizing martingale is the Doob-Meyer martingale component of the option price process, namely the "optimal dual martingale", then strong duality is achieved, meaning that the upper bound is tight. [69] and [24] generalize the idea of duality theory from American option pricing to general discrete-time DPs and provide a broader interpretation of the dual martingale. From the perspective of [24], the dual martingale could be regarded as the penalty for the access to the outcomes of future asset prices. Furthermore, similar to the existence of an optimal dual martingale in option pricing, [24] shows that in a general DP there also exists an optimal dual penalty such that the resulted dual problem recovers the optimal value function. In particular, one specific choice of optimal dual penalties is the Doob-Meyer martingale component of the optimal value function process with respect to (w.r.t.) the natural information filtration at terminal horizon (referred to as the value-based optimal dual penalty).

The idea of relaxing the non-anticipativity constraint on the policies of DPs at least dates back to [31], [32]. Moreover, it has been extensively applied in solving stochastic programs, in which the dualization results from the Lagrangian relaxation of the non-anticipativity constraint on the decision variables, see [78], etc.

There have been many new methodologies and applications involving the duality theory of DP in recent years. [27] studies the additive and multiplicative dual formulations of American option pricing problems. [23] and [22] propose the gradient/subgradient penalties for the dual problems of convex DPs (concave reward function and convex action sets), in order to preserve the convexity structure in the dual problems and facilitate the optimization. The authors further show that these penalties could be optimal under appropriate conditions. [86] generalizes the duality theory to controlled Markov diffusion (CMD) under a continuous-time setting and

reveal the structure of the optimal dual penalty as a stochastic integral. [20] generalizes the duality theory to discrete-time infinite-horizon MDPs and propose several techniques to approximate the optimal dual penalty. As for the applications, [56] and [57] apply duality theory to generate upper bounds in the valuation of natural gas storage. [43], [41], [62], and [30] apply the duality theory to measure the performances of different trading strategies in portfolio optimization problems. [45] studies the closed-form expression of the optimal dual penalty in linear quadratic control. [21] and [85] apply duality theory to multi-armed bandits problem and weakly coupled dynamic programs, respectively. [52] studies a robust multi-armed bandits problem using the information relaxation approach. [42] applies duality theory to study zero-sum games.

We also note that there are many new developments in computational methods that aim at constructing good dual penalties. In general, the value-based optimal dual penalty could not be computed exactly because it involves optimal value functions that are not available and conditional expectations that need to estimated. The naive approach is to replace the optimal value functions with approximate ones, and use nested simulation to estimate the conditional expectations; however, this approach often requires substantial computational effort. Various methods have been proposed to approximate the optimal dual penalty accurately, or generate good dual penalties efficiently. [12] and [89] propose the non-nested simulation approach in American-style option pricing under diffusion processes and jump-diffusion processes, respectively, by revealing the structure of the optimal dual martingales as stochastic integrals. They further develop efficient schemes to fast approximate those integrals. [36] proposes to use parameterized martingales to improve the quality of the upper bounds for optimal stopping problems by solving a convex optimization problem. [35] and [84] further generalize this idea to generate parameterized dual penalties for general DPs via convex pathwise optimization. The advantage of this approach is

that it explores a subspace of feasible dual penalties by considering the best linear combination of the existing dual penalties. However, the drawback is that it requires solving a new stochastic optimization problem, which could be computationally expensive or intractable. Moreover, the quality of the resulted dual penalty heavily relies on the quality of the existing dual penalties; hence, good performance could not be guaranteed.

1.1.2 Motivation and Research Goals

We notice that two key things are missing in most of the existing approaches:

- The structure of the optimal dual penalty has not been not well-studied. Since the space consisting of all feasible dual penalties is a function space (referred to as the dual penalty space hereafter), the optimal dual penalty could be viewed as a point in that space. Therefore, we could approximate it by computing and estimating all its coordinates w.r.t. a (properly chosen) functional basis of the dual penalty space. If such estimation could be achieved without Monte Carlo simulation, then nested simulation is circumvented.

- The dual problem is usually solved independently of the primal problem, meaning that some useful information (e.g., the suboptimal policy has been evaluated and the simulations have been carried out) in the primal problem is not well-utilized. Properly utilizing that information might facilitate the estimation of coordinates.

Motivated by these observations, one of our research goals in this thesis is to design a scheme and approximate the optimal dual penalty accurately and efficiently for general DPs. In particular, we are interested in exploring the structure of the dual penalty space as well as the optimal dual penalty through constructing good functional bases. Therefore, one could approximate the optimal dual penalty by

estimating its coordinates w.r.t. the functional basis. The questions we attempt to answer include:

- Is the dual penalty space a well-defined function space? If so, how to construct proper functional bases of that space such that the structure of the optimal dual penalty could be understood through studying the corresponding coordinates? In particular, what properties should a function basis possess to facilitate the computation and estimation of coordinates?

- How to estimate the coordinates efficiently with minimal extra simulation and computational costs? In particular, is it possible to reutilize the information such as the suboptimal policy and the sample paths that were generated in the primal problem to estimate the coordinates? Does it incur nested Monte Carlo simulation during this procedure?

- Suppose the coordinates of the optimal dual penalty have been estimated, which results in an approximation of the optimal dual penalty. Is it a feasible dual penalty?

- Is this approximation scheme sufficiently robust to generate dual penalties with desired structural properties? For example, for convex DPs (concave reward functions and convex constraints on actions), is this scheme capable of generating dual penalties that preserve convexity in the dual problem?

1.1.3 Research Results and Contributions

To answer all the aforementioned research questions, we propose a framework of regression approach that explores the structure of the dual penalty space as well as the optimal dual penalty while utilizing the information from the primal problem. In particular, we will show that the dual penalty space is a well-defined function space, then we will construct functional bases of that space such that the corresponding

coordinates of the optimal dual penalty are easy to compute by solving systems of linear equations. Furthermore, instead of using nested Monte Carlo simulation to estimate those coordinates, we propose an efficient regression method that reutilizes the information from the primal problem.

The proposed framework has the following advantages. Firstly, it is a scheme that requires minimal extra simulation or computational costs since it only requires using the suboptimal policy and sample paths that were generated in the primal problem. Secondly, it generates approximations of the optimal dual penalty without nested simulation, which is suffered by some existing approaches; therefore, computational efficiency is significantly improved. Thirdly, the resulted dual penalties are feasible, and thus they induce valid upper bounds on the optimal value function. Finally, as we will demonstrate later, the framework is very robust; dual penalties with desired structural properties could be generated by constructing proper regressors in the regression.

In a broader sense, several existing approaches to approximating the optimal dual penalty could be regarded as special cases of this framework under specific settings. For instance, [12] proposes a non-nested simulation approach to approximating the optimal dual martingale in American option pricing under diffusion processes. We will show that their approach can be viewed as a special case of the proposed framework with a specific functional basis of the dual penalty space. Our framework is more universal and powerful, because it reveals the structure of the optimal dual penalty regardless of the underlying probability measure (i.e., not restricted to the Brownian measure in [12]). We will show that the approximation scheme in [86] can be viewed as a special case of the proposed framework as well. To summarize, the contributions are as follows:

- We develop a framework of regression approach to approximating the optimal dual penalty for a general DP that (1) circumvents nested simulation suffered

by some existing approaches, (2) requires minimal extra simulation or computational costs, (3) yields feasible dual penalties and valid upper bounds, and (4) capable of generating dual penalties with desired structural properties by properly constructing the regressors in the regression.

- We explore the dual penalty space and reveal the structure of the optimal dual penalty for general DPs, which enable us to generate good feasible dual penalties in a systematic way. Several existing approaches could be regarded as special cases of the framework under specific settings.

- The application of the proposed framework to a high-dimensional dynamic trading problem demonstrates its effectiveness and efficiency in solving the duals of complex DPs. In particular, it generates accurate approximations of the optimal dual penalty and tight upper bounds on the optimal value function when good suboptimal policies and appropriate regressors are used in the regression approach.

1.2 Assessing Risk in Stochastic Simulation under Input Uncertainty

For a complex real-world stochastic system, simulation is a powerful tool to analyze its behavior when real experiments on the system are expensive or difficult to conduct. Simulation is driven by input models that are distributions capturing the randomness in the system. For example, in simulating a queueing network, random customer arrivals and service times need to be generated from appropriate distributions (i.e., input models). The uncertainty on the input parameters (e.g., customer arrival rates and mean service times for different customer types) may need to be taken into account, since they are typically estimated from finite records of previous data. In general, there are two sources of uncertainty for a typical stochastic simulation experiment: the extrinsic uncertainty on input parameters (referred to as *input*

parameter uncertainty, or simply *input uncertainty*) that reflects variability of the finite data used to estimate input parameters, and the intrinsic uncertainty on output response (referred to as *stochastic uncertainty*) that reflects the inherent stochasticity in the system. There might be also a third source of uncertainty—*model uncertainty*, which refers to the uncertainty of the distribution family used in the input models.

The variability of simulation output response clearly depends on both stochastic uncertainty and input uncertainty. An important question to address is how to quantify the impact of input uncertainty on output response variability. Various quantification methods have been proposed, including frequentist and Bayesian methods among many others. Frequentist methods include the Direct/Bootstrap Resampling methods by [8], [9], [28], etc. The input model for these methods can be a non-parametric empirical distribution or a parametric distribution estimated from historical data. Bayesian methods include the Bayesian Model Averaging (BMA) methods by [29], [92], [93], [15], etc. In these methods, the Bayesian updating rule is applied on a chosen prior distribution of input parameters to obtain a posterior parameter distribution, which is then used as the sampling distribution of input parameters in the simulation experiment. In addition to the aforementioned methods, [28] also develops the δ-method, which is based on Taylor's Theorem to decompose the variance of simulation output response into two components that are caused by input parameter uncertainty and stochastic uncertainty, respectively. [46] provides an early review on the importance of considering input uncertainty and common methods to deal with it. [6] provides a more recent review on popular methods in output analysis with input uncertainty and highlighted some remaining challenges in this area. In recent years, with the rise of stochastic kriging meta-modeling method in stochastic simulation (e.g., [4]), meta-model assisted methods have been developed for quantifying input uncertainty, see [7], [82], [83], etc.

Many application of these methods aim at providing inferences on the average

behavior of mean response under all possible input models, often through a point estimation and its associated confidence interval (CI). Some others focus on obtaining an empirical distribution of mean response, providing a more complete picture of all possible behaviors of mean response under input uncertainty.

1.2.1 Motivation and Research Goals

Most of the aforementioned works aim at providing inferences on the average behavior of mean response under all possible input models. However, a rigorous quantification of extreme behavior of mean response in all possible input models is still lacking. Such quantification could provide inferences on system sensitivity or stability, and thus would be critical for risk assessment and control of the system — one of the fundamental factors in real-world system design.

For example, consider the system of a typical hospital emergency room (ER). When the administrators of ER determine the number of on-call doctors, one of the main criteria is the expected number of waiting patients or the expected patient waiting time. Quantifying and controlling the extreme behavior of mean response (e.g., the expected number of waiting patients) in all possible input models are quite necessary, because an extreme mean response indicates a large number of expected waiting patients, which might lead to delayed treatment of patients and possibly serious consequences in life-threatening situations.

For another example, consider a large-scale power system. It is usually too expensive or risky to conduct real experiments on the system operation; therefore, stochastic simulation is often used to study the economics, reliability, and emission variable effects of power systems operating in a market environment (see, e.g., [34]). In a typical power system simulation experiment, the inputs may include the resource parameters, the loading (market demand) parameters, etc., which all exhibit variability and uncertainty. The risk assessment and control of system performance under

input uncertainty is of great importance because the extreme behavior of mean response (e.g., the expected power balance in peak time) might cause a part or whole breakdown of the power system and lead to disastrous outcomes.

It motivates us to study the assessment of the risk of extreme mean responses due to input uncertainty by studying risk measures of mean response w.r.t. the distribution of input parameter. In particular, we will consider the following two mostly common used risk measures: Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Loosely speaking, VaR characterizes the extreme (e.g., 99%) quantile of the mean response distribution, and CVaR characterizes the conditional mean of the very tail portion of the mean response distribution. VaR, as one of the very earliest risk measures introduced in financial risk management, is easy to understand and interpret for practitioners. CVaR, as a classic coherent risk measure (see, e.g., [5]), exhibits nice properties such as convexity and monotonicity for optimization (see, e.g., [67]). They have been extensively used in financial society, especially after the financial crisis in 2008. There is abundant existing literature that dedicates to studying the estimation and optimization of risk measures under various settings; in particular, [48] provides an elegant review of Monte Carlo methods for VaR and CVaR.

We are motivated by following research questions:

- How to properly define risk (VaR and CVaR) of stochastic simulation under input uncertainty? Furthermore, how to construct well-behaved estimators for those risk measures under input uncertainty? Is nested simulation involved?

- Suppose the VaR and CVaR estimators are nested. What properties do they possess? In general, one would care about unbiasedness, consistency and asymptotic normality of an estimator. Then how to properly characterize these properties for estimators that involve nested simulation?

- Is it possible to construct CIs associated with the nested point estimators of risk measures under input uncertainty? Are they valid CIs when the number of samples goes to infinity?

- Suppose the estimators involve a two-layer nested simulation. Intuitively, the outer-layer simulation captures input uncertainty and the inner-layer simulation captures stochastic uncertainty. Further assume the total simulation budget (sample size) is fixed. How to smartly allocate the budget between inner simulation and outer simulation, and balance between capturing input uncertainty and capturing stochastic uncertainty?

1.2.2 Research Results and Contributions

We will introduce risk measures such as VaR and CVaR to quantify the risk of stochastic simulation under input uncertainty, and provide numerical schemes for their estimation. Specifically, we will study nested Monte Carlo estimators for VaR and CVaR of mean response from both theoretical and computational aspects. In particular, we will show some non-trivial theoretical results regarding the important properties (unbiasedness, consistency and asymptotic normality) of the proposed nested risk estimators. Last, we will present numerical examples to illustrate the importance and necessity of risk assessment in a stochastic system under input uncertainty. To summarize, the contributions are three-folds:

- For output analysis in stochastic simulation, this work is among the first to systematically study risk measures that rigorously quantify the extreme behavior of mean response in all possible input models.

- Under the respective "Weak Assumption" and "Strong Assumption", which are elaborated later, we show that nested risk estimators are strongly consistent in weak and strong limiting sense, respectively. We further show the asymptotic

normality of nested risk estimators, which is the guarantee for constructing (asymptotically valid) CIs.

- A novel approach is developed to solve the associated budget allocation problem that arises in nested simulation of risk estimators, in order to improve simulation efficiency. The numerical study demonstrates its effectiveness by showing that the obtained budget allocation schemes drastically reduce the CI widths of risk estimators.

We note that the risk formulation of input certainty bears great similarity with risk assessment in portfolio/credit management. In a broader sense, they both deal with simulating certain conditional expectations. The work most relevant to ours is probably [40], which studies the asymptotic representation of Mean Square Error (MSE) of nested risk estimators in credit risk management. By minimizing MSE asymptotically, the authors obtain an (asymptotically) optimal budget allocation scheme. In contrast, our work focus on the strong consistency and asymptotic normality of nested risk estimators. Furthermore, the associated budget allocation problem in our approach is to minimize the CI width of nested risk estimators, and hence our solution strategy and resulting optimal budget allocation scheme are drastically different from [40]'s. Nevertheless, part of our analysis follows from the assumptions and analysis in [40]. Other common approaches for risk assessment in portfolio/credit management include but not limited to the delta-gamma method by [72], [39], etc; the two-level CI procedure with screening by [58], etc; the stochastic kriging method by [61], etc; the ranking and selection method by [18], etc; the regression method by [19], etc. Among other relevant literature, [60] studies the point estimation of a quantile (VaR) of the distribution of a conditional expectation via two-level simulation; [79] considers estimating the density of a conditional expectation using kernel density estimation; [81]

studies efficient nested simulation for estimating the variance of a conditional expectation. Most of these works focus on efficient budget allocation of inner samples for different outer scenarios in nested estimation, and [60], [79], and [81] consider optimal allocation between inner and outer sampling. Our work is distinct from these works in that we focus on the theoretical properties (strong consistency and asymptotic normality) of nested risk estimators, and our budget allocation can be viewed as a byproduct of the asymptotic normality results that we have established. Of course, optimal budget allocation among inner samples, as done in some of these works, can be incorporated to further improve our nested risk estimators; however, this is not our focus in this proposal.

1.3 Optimizing Risk Measures via Gradient-based Stochastic Adaptive Search

As mentioned previously, risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are widely studied in various fields to quantify the extreme behaviors of functions of interest. They have been extensively used in the financial industry, especially after the financial crisis in 2008. An abundant literature has dedicated to studying the estimation and optimization of risk measures under various settings. To list a few, [67] and [68] derive some fundamental properties of CVaR for general loss distributions in finance, and propose the fundamental minimization formula to facilitate the optimization of CVaR. [75] develop the dual theory for optimization of general coherent risk measures, and derive the optimality conditions via the dual representation. [74] study the optimization of risk measures under a multistage setting, and propose a risk-averse dynamic programming approach to Markov decision processes. [2] study the optimization of VaR and CVaR minimization for derivatives portfolios with the addition of a cost that is proportional to the portfolio position. [47] and [49] study the efficient Monte Carlo estimation of the VaR derivative and CVaR derivative, respectively.

In general, optimizing risk measures over continuous decision variables is a challenging problem, especially when the underlying loss function does not possess good structural properties such as convexity or differentiability. Traditional gradient-based optimization methods often are not applicable, since little problem-specific knowledge is known when the loss function is evaluated via black-box simulation of a stochastic system. In contrast, model-based optimization methods are good alternatives as they impose minimal assumptions on the problem structure, which include ant colony optimization ([37]), annealing adaptive search (AAS) ([71]), the estimation of distribution algorithms (EDA) ([59]), the cross-entropy (CE) method ([73]), model reference adaptive search (MRAS) ([50] and [51]), the interacting-particle algorithm ([63] and [64]), gradient-based adaptive stochastic search (GASS) ([87]), etc.

The main idea of model-based methods is to introduce a sampling distribution, which often belongs to a parameterized family of densities, over the solution space, and iteratively update the sampling distribution (or its parameter) by drawing and evaluating candidate solutions according to the sampling distribution. The hope is to have the sampling distribution more and more concentrated on the promising region of the solution space where the optimal solutions are located, and eventually become a degenerate distribution on one of the global optima. Therefore, finding an optimal solution in the solution space is transformed to finding an optimal sampling distribution parameter in the parameter space. A key difference among the aforementioned model-based methods lies in how to update the sampling distribution. For example, in MRAS and GASS, the updating rule on the sampling distribution parameter is derived by converting the original (possibly non-differentiable) deterministic optimization problem into a differentiable stochastic optimization problem on the sampling distribution parameter, and then applying stochastic approximation schemes. Compared with gradient-based methods, model-based methods are more robust in the sense that at every iteration they exploit the promising region of the

solution space that has already been identified, while maintaining the exploration of the entire solution space. The updating rule on the sampling distribution parameter controls the balance between the exploration and the exploitation.

1.3.1 Motivation and Research Results

Although all the aforementioned model-based methods are designed for deterministic optimization problems, they can be extended to risk (VaR or CVaR) optimization problems in which the exact risk values are replaced with sample estimates (though might be biased). However, a straightforward extension usually leads to an algorithm that is computationally expensive, due to the rare-event simulation required in estimating the risk values. This issue is even more severe when a large risk level (close to 1) is of interest. It motivates us to consider the following question: is it possible to initialize a model-based algorithm for a risk optimization problem with a small risk level (close to 0), and then adaptively adjust or increase the risk level at every iteration such that the target risk level is achieved and the algorithm converges simultaneously? The hope is that the algorithm will consume less simulation budget (since the risk level is small) during the "warm-up" phase of the algorithm, solve problems that are close to the original one during the "convergence" phase of the algorithm, and eventually achieve budget saving. The key to this question lies in finding a signal to link the updating rule on the risk level with the updating rule on the sampling distribution parameter, where a signal is capable of measuring empirically the algorithm's emphasis between the exploitation of a promising region and the exploration of the entire solution space.

In this thesis, we will focus on the extension of a specific model-based methods—GASS by [87] to the optimization of risk measures. We choose GASS because it can also be interpreted as a gradient-based scheme of a reformulated problem, in which a Newton-like updating rule is applied on the sampling distribution parameter, and

thus the gradient (even the Hessian) in the updating rule of the sampling distribution parameter can be viewed as a signal that empirically measures the algorithm's emphasis between the exploitation of a promising region and the exploration of the entire solution space. Therefore, we could adjust the risk level adaptively using the information contained in the gradient (e.g., its norm) at every iteration. In particular, we will propose an updating rule that increases the risk level proportionally to the decrease in the norm of the gradient. To the best of our knowledge, this work is among the first to apply model-based algorithms to risk optimization problems, and among the first to propose a risk optimization scheme that adaptively adjusts the risk level. For the ease of presentation, we will only focus on CVaR optimization, and the extension of the proposed algorithm to VaR optimization (and possibly other risk measures such as probability of large loss) is straightforward.

1.4 Outlines

The next three chapters provide the details on how we conduct the proposed methodologies on the research problems, and they are organized as follows.

In Chapter 2, we discuss in details the proposed framework of regression approach for solving the dual problems of general DPs. The application to a high-dimensional dynamic trading problem with predictable returns and transaction costs is also presented to show its effectiveness in solving dual problems of complex DPs. In Chapter 3, we discuss in details the assessment of risk of stochastic simulation under input uncertainty. In particular, we show some non-trivial results regarding the properties of the proposed nested risk estimators. In Chapter 4, we discuss in details the optimization of risk via GASS with adaptive risk adjustments.

Extra numerical results for the dynamic trading problem studied in Chapter 2 are presented in Appendix A. Proofs of the theorems in Chapter 3 are presented in Appendix B. Proofs of the theorems in Chapter 4 are presented in Appendix C.

CHAPTER II

SOLVING THE DUALS OF DYNAMIC PROGRAMS VIA REGRESSION

In this chapter we develop the framework of regression approach to approximating the optimal dual penalties of general dynamic programs (DP). It is also presented in our paper [90].

The rest of this chapter is organized as follows. In Chapter 2.1, we review the basics of dynamic programming and information relaxation duality theory. We present the framework of regression approach in Chapter 2.2. In Chapter 2.3, we show several existing approaches are special cases of the proposed framework under specific settings. In Chapter 2.4, we apply the proposed framework to a high-dimensional dynamic trading problem, and demonstrate its effectiveness and efficiency in solving the duals of complex DPs. Conclusion is provided in Chapter 2.5.

2.1 Dynamic Programs and Dual Formulations

On a general probability space $(\Omega, \mathbb{P}, \mathscr{F})$, in which Ω is the set of all possible outcomes (scenarios) of uncertainties, $\mathbb{P}$ is the underlying probability measure and $\mathscr{F}$ is the σ-algebra consisting of all the events (measurable subsets of Ω), consider a finite-horizon MDP as follows. Time is indexed by $\mathcal{T} = \{0, 1, ..., N\}$. The state x follows the dynamics

$$x_{n+1} = f(x_n, a_n, z_{n+1}), \quad n = 0, 1, \cdots, N-1, \tag{2.1.1}$$

where f is the deterministic transition function, $x_n \in \mathscr{X}_n$ denotes the state at period n that lives in the state space $\mathscr{X}_n$, $a_n \in \mathscr{A}_n$ denotes the action/control at period n that is chosen from the action space $\mathscr{A}_n$, z_{n+1} is the random noise at period n and

$\{z_n : n = 1, ..., N\}$ are assumed to be independently and identically distributed (i.i.d.) random variables with probability measure ρ on support space $\Xi \subseteq \mathbb{R}^d$. The evolution of information is described by the natural filtration $\mathcal{F} = \{\mathscr{F}_n : n = 0, ..., N\}$. Loosely speaking, $\mathscr{F}_n$ describes the information available to the DM at period n. In particular, each z_n is $\mathscr{F}_n$-measurable. Without loss of generality, we further assume $\mathscr{F}_0 = \{\emptyset, \Omega\}$, meaning that the DM initially does not have any knowledge about the outcomes of uncertainties, and $\mathscr{F}_N = \mathscr{F}$, meaning that the DM knows all the possible outcomes of uncertainties in the end.

We use a mapping $\alpha_n(\cdot)$ from the state space $\mathscr{X}_n$ to the action space $\mathscr{A}_n$, i.e., $\alpha_n : \mathscr{X}_n \to \mathscr{A}_n$, to denote the decision rule at period n. A policy/strategy $\boldsymbol{\alpha} := (\alpha_0, \alpha_1, ..., \alpha_{N-1})$ consisting of a sequence of decision rules is called non-anticipative/$\mathcal{F}$-adapted, if each decision rule $\alpha_n(\cdot)$ is $\mathscr{F}_n$-measurable. Intuitively, it means the DM chooses the action a_n only based on the information accumulated up to period n, he/she shall not choose the action a_n based on the future information. We use $\mathbb{A}_{\mathcal{F}}$ to denote the set of all non-anticipative policies, and $\mathbb{A}$ to denote the set of all policies (including the anticipative ones); clearly $\mathbb{A}_{\mathcal{F}} \subseteq \mathbb{A}$. Furthermore, we associate an $\mathscr{F}_n$-measurable reward function $r_n(x_n, a_n)$ with the state dynamics to represent the immediate reward after the DM chooses action a_n at period $n = 0, ...N-1$, and an $\mathscr{F}_N$-measurable function $r_N(x_N)$ as the terminal reward at period N.

Given $x_0 \in \mathscr{X}_0$, the objective of the DM is to select a non-anticipative policy $\boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}$ that maximizes the accumulated rewards over all the periods, i.e.,

$$(P): \ V_0(x_0) \overset{\triangle}{=} \sup_{\boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}} V_0^{\boldsymbol{\alpha}}(x_0) = \sup_{\boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}} \mathbb{E}_0 \left[\sum_{n=0}^{N-1} r_n(x_n, \alpha_n(x_n)) + r_N(x_N) \right], \qquad (2.1.2)$$

where $\mathbb{E}_0[\cdot]$ means that the expectation is taken w.r.t. $\mathscr{F}_0$, and we will use $\mathbb{E}_n[\cdot]$ to denote the expectation taken w.r.t. $\mathscr{F}_n$ thereafter.

It is well known that problem (2.1.2) could be recursively solved theoretically via

Bellman backward dynamic programming

$$\begin{cases} V_N(x_N) \triangleq r_N(x_N), \\ V_n(x_n) \triangleq \sup\limits_{a_n \in \mathscr{A}_n} \{r_n(x_n, a_n) + \mathbb{E}_n[V_{n+1}(x_{n+1})]\}, \quad n = N-1, ..., 0, \end{cases} \tag{2.1.3}$$

where $V_n(x_n)$ represents the optimal value function of the DP with initial period n and initial state x_n. However, in practice, the Bellman recursion (2.1.3) could hardly be solved exactly for most cases, due to the curse of dimensionality. Therefore, one often needs to settle with suboptimal policies or approximate optimal value functions. A suboptimal policy could be evaluated conveniently via Monte Carlo simulation to generate a lower bound on the optimal value function. However, in the absence of the exact optimal value function or its upper bounds, the quality of the suboptimal policy could hardly be measured.

2.1.1 A Dual Formulation

The duality theory developed by [69] and [24] addresses the aforementioned issue by formulating a dual representation of the (primal) DP (2.1.2), and solving the dual problem provides a valid upper bound on the optimal value function. Therefore, the quality of a suboptimal policy could be empirically measured by examining the duality gap. If it is sufficiently small, then the suboptimal policy could be claimed to be near optimal. To be more specific, let us rigorously define a *feasible dual penalty* as follows.

Definition 2.1.1. *We say $M(\boldsymbol{\alpha}, \mathbf{z})$, a functional of policy $\boldsymbol{\alpha} \in \mathbb{A}$ and noise sequence $\mathbf{z} := (z_1, ..., z_N)$, is a feasible dual penalty if*

$$\mathbb{E}_0[M(\boldsymbol{\alpha}, \mathbf{z})] = 0, \quad \forall \boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}. \tag{2.1.4}$$

Put in another way, a penalty function $M(\boldsymbol{\alpha}, \mathbf{z})$ is dual feasible if it does not penalize any non-anticipative policy in expectation. We further use $\mathbb{M}_{\mathcal{F}}$ to denote the set of all feasible dual penalties.

Remark 2.1.1. *Definition 2.1.1 is slightly different from the one in [24], in which a dual penalty is called feasible if* $\mathbb{E}_0\left[M(\boldsymbol{\alpha}, \mathbf{z})\right] \leq 0,\ \forall \boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}$. *The reason for using Definition 2.1.1 is to ensure that the set of all feasible dual penalties* $\mathbb{M}_{\mathcal{F}}$ *is a function space (vector space). Note that Definition 2.1.1 does not exclude any "good" feasible dual penalties in [24], since* $(M(\boldsymbol{\alpha}, \mathbf{z}) - \mathbb{E}_0\left[M(\boldsymbol{\alpha}, \mathbf{z})\right]) \in \mathbb{M}_{\mathcal{F}}$ *always induces an upper bound as tight as the one induced by* $M(\boldsymbol{\alpha}, \mathbf{z})$. *In particular, it does not exclude the optimal dual penalty (which will be defined in the following).*

[24] shows that the following dual representation of the primal DP (2.1.2), derived by subtracting a feasible dual penalty $M(\boldsymbol{\alpha}, \mathbf{z}) \in \mathbb{M}_{\mathcal{F}}$ from the objective function and relax the non-anticipativity constraint on all the feasible policies, i.e.,

$$\begin{aligned}(D): \quad V_0^M(x_0) &= \sup_{\boldsymbol{\alpha}\in\mathbb{A}} \mathbb{E}_0\left[\left\{\sum_{n=0}^{N-1} r_n\left(x_n, \alpha_n(x_n)\right) + r_N(x_N) - M(\boldsymbol{\alpha}, \mathbf{z})\right\}\right] \\ &= \mathbb{E}_0\left[\sup_{\mathbf{a}\in\mathscr{A}}\left\{\sum_{n=0}^{N-1} r_n\left(x_n, a_n\right) + r_N(x_N) - M(\boldsymbol{a}, \mathbf{z})\right\}\right], \quad (2.1.5)\end{aligned}$$

yields an upper bound $V_0^M(x_0)$ on the optimal value function $V_0(x_0)$, where $\mathbf{a} := (a_0, ..., a_{N-1})$ represents the action sequence and $\mathscr{A} := (\mathscr{A}_0, ..., \mathscr{A}_{N-1})$. Conceptually, the dual problem (2.1.5) consists of a series of deterministic pathwise optimization problems (referred to as inner optimization problems), in which the actions could be chosen with full knowledge of all outcomes of uncertainties. In practice, solving dual problem (2.1.5) is convenient via Monte Carlo simulation given that the inner optimization problems are tractable: simulate multiple i.i.d. noise sequences $\mathbf{z}$, then solve the deterministic inner optimization problem corresponding to each $\mathbf{z}$, and finally take the average of the optimal values as an estimator of $V_0^M(x_0)$.

Note that a dual penalty is feasible if and only if it does not penalize any non-anticipative policy. We immediately have the following lemma on weak duality.

Lemma 2.1.1. *(**Weak Duality**) [Lemma 2.1 in [24]] If* $M(\boldsymbol{\alpha}, \mathbf{z}) \in \mathbb{M}_{\mathcal{F}}$, *then we have*

$$V_0(x_0) \leq V_0^M(x_0).$$

Lemma 2.1.1 suggests that the dual problem (2.1.5) with any feasible dual penalty could be used to generate an upper bound on the optimal value function. The next question is whether there exists an optimal dual penalty that closes the duality gap and recovers the optimal value function, i.e., the existence of strong duality. It turns out, if we minimize $V_0^M(x_0)$ over all feasible dual penalties $M \in \mathbb{M}_{\mathcal{F}}$, then we have $\forall x_0 \in \mathscr{X}_0$,

$$\sup_{\boldsymbol{\alpha}\in\mathbb{A}_{\mathcal{F}}} V_0^{\boldsymbol{\alpha}}(x_0) = V_0(x_0) = \inf_{M\in\mathbb{M}_{\mathcal{F}}} V_0^M(x_0),$$

i.e., strong duality holds. In particular, the following theorem shows that the value-based optimal dual penalty in the form of (2.1.6) suffices for the strong duality to hold.

Theorem 2.1.1. *(**Strong Duality**) [Theorem 2.3 in [24]] Let $\{V_n(x_n)\}$ be the optimal value functions of the primal DP (2.1.2). Further let $M^*(\boldsymbol{\alpha}, \mathbf{z})$ be the martingale difference sum of $\{V_n(x_n)\}$, i.e.,*

$$M^*(\boldsymbol{\alpha}, \mathbf{z}) \triangleq \sum_{n=0}^{N-1} \left(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1})|x_n, a_n\right]\right), \quad \text{where } a_n = \alpha_n(x_n), \quad (2.1.6)$$

then $M^*(\boldsymbol{\alpha}, \mathbf{z})$ is a feasible dual penalty. Moreover, it achieves strong duality, i.e.,

$$V_0^{M^*}(x_0) \triangleq \mathbb{E}_0\left[\sup_{\boldsymbol{\alpha}\in\mathbb{A}}\left\{\sum_{n=0}^{N-1} r_n\left(x_n, \alpha_n(x_n)\right) + r_N(x_N) - M^*(\boldsymbol{\alpha}, \mathbf{z})\right\}\right] = V_0(x_0). \quad (2.1.7)$$

Note that the optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z})$ defined in (2.1.6) is a functional of $(\boldsymbol{\alpha}, \mathbf{z})$ because the state sequence $(x_0, ..., x_N)$ and action sequence $\mathbf{a}$ depend on $(\boldsymbol{\alpha}, \mathbf{z})$ through the state dynamics (2.1.1) and the decision rule mappings. Intuitively, Theorem 2.1.1 implies that the advantage gained by the access to the future information is perfectly cancelled out in expectation by the optimal dual penalty. What is more striking about the optimal dual penalty is that the second equality in (2.1.7) is achieved almost surely for every inner optimization problem scenario. Therefore, we can drop the expectation sign in (2.1.7) and strengthen the result as

$$V_0^{M^*}(x_0) = \sup_{\boldsymbol{\alpha}\in\mathbb{A}}\left\{\sum_{n=0}^{N-1} r_n\left(x_n, \alpha_n(x_n)\right) + r_N(x_N) - M^*(\boldsymbol{\alpha}, \mathbf{z})\right\} = V_0(x_0), \quad a.s..$$

It implies that, in practical implementation, if the approximation of the optimal dual penalty is sufficiently accurate, then the variance of an one-sample estimator of the upper bound is small. Therefore, one only need to solve a small number of deterministic inner optimization problems to generate a good upper bound estimator.

In theory, strong duality implies that the upper bound induced by the optimal dual penalty is tight. However, one could hardly compute the optimal dual penalty exactly and close the duality gap, since in general the optimal value functions $\{V_n(x_n)\}$ are not available. A naive alternative is to replace the optimal value functions with approximate value functions $\{\widetilde{V}_n(x_n)\}$ that might be induced by a suboptimal policy or certain approximate dynamic programming technique. The computational effort required in approximating the optimal dual penalty should be taken into consideration as well. A common method to estimate the conditional expectations in the optimal dual penalty (2.1.6) is nested simulation, which generates scenarios in the outer-layer and uses sample averaging in the inner-layer associated with each scenario. Hence, the total simulation effort is proportional to the number of outer-layer scenarios multiplied by the number of inner-layer samples. Albeit stable with a large-scale simulation, it might not be a desirable choice facing a limited computational budget. Several approaches have been developed in recent years that aim to generate good dual penalties without nested simulation. However, we note that two key things are missing in most of the existing approaches. The structure of the dual penalty space as well as the optimal dual penalty is not well-studied, and the information such as the suboptimal policy and the sample paths generated in solving the primal problem is not well-utilized.

In the next subsection, we will present a framework of regression approach to approximating the optimal dual penalty that explores the structure of dual penalty space and efficiently reutilizes the information in the primal problem.

2.2 A Regression Approach

2.2.1 General Framework

The main idea of the regression approach is to view the optimal dual penalty as a point in the dual penalty space, then compute its coordinates w.r.t. a properly chosen functional basis, and finally estimate those coordinates in a non-nested manner. The key question lies in how to choose the functional basis and compute the corresponding coordinates of the optimal dual penalty. Let us first derive the general framework. Recall the optimal dual penalty

$$M^*(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{n=0}^{N-1} \left(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n \right] \right), \quad \text{where } a_n = \alpha_n(x_n). \tag{2.2.1}$$

Denote the n-th single martingale difference term in $M^*(\boldsymbol{\alpha}, \mathbf{z})$ by $M_n^*(\boldsymbol{\alpha}, \mathbf{z})$, i.e.,

$$M_n^*(\boldsymbol{\alpha}, \mathbf{z}) \triangleq \left(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n \right] \right), \ n = 0, ..., N-1. \tag{2.2.2}$$

Then clearly $M^* = \sum_{n=0}^{N-1} M_n^*$. Further let $\mathcal{B} = \{b_i : i \in I\}$ denote a functional basis of $\mathbb{M}_{\mathcal{F}}$, where $b_i(\cdot)$ is a function on the support space Ξ of the noises $\{z_n\}$. Note that $\mathcal{B}$'s cardinality $|I|$ might be countable or uncountable. We will discuss about the selection of $\mathcal{B}$ or the properties it should possess later. Suppose we could express $M_n^*(\boldsymbol{\alpha}, \mathbf{z})$ w.r.t. $\mathcal{B}$ in format as

$$\left(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n \right] \right) = \sum_{i \in I} \beta_{n,i}(x_n, a_n) \cdot b_i(z_{n+1}), \ n = 0, ..., N-1, \tag{2.2.3}$$

where $\{\beta_{n,i}(x_n, a_n)\}$ are the coordinates. To compute $\{\beta_{n,i}(x_n, a_n)\}$, multiplying $b_i(z_{n+1})$ $\forall i \in I$ on both sides of (2.2.3) and taking conditional expectations w.r.t. $\mathscr{F}_n$, we obtain

$$\begin{aligned} &\mathbb{E}_n\left[V_{n+1}(x_{n+1}) \cdot (b_i(z_{n+1}) - \mathbb{E}_n[b_i(z_{n+1})]) \right] \\ = &\sum_{j \in I} \beta_{n,j}(x_n, a_n) \mathbb{E}_n\left[b_i(z_{n+1}) b_j(z_{n+1}) \right], \end{aligned} \tag{2.2.4}$$

where note that $\mathbb{E}_n[b_i(z_{n+1})]$ and $\mathbb{E}_n\left[b_i(z_{n+1}) b_j(z_{n+1}) \right]$ are constants that only depend on $\mathcal{B}$. Therefore, in principle we could view (2.2.4) as a system of linear equations

with variables $\{\beta_{n,i}(x_n, a_n)\}$ (although the number of equations in the system could be countably infinite or even uncountable). Assuming it is solvable (again this depends on the choice of $\mathcal{B}$), then we have

$$\beta_{n,i}(x_n, a_n) = \mathbb{E}\left[V_{n+1}(x_{n+1}) \cdot h_i(z_{n+1})|x_n, a_n\right],\ i = 1, ...,\ n = 0, ..., N-1, \quad (2.2.5)$$

where $\{h_i(\cdot)\}$ are deterministic functions that only depend on $\mathcal{B}$ because all the parameters of the linear system are uniquely determined by $\mathcal{B}$. It follows that the optimal dual penalty could be rewritten as

$$M^*(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{n=0}^{N-1} \sum_{i \in I} \beta_{n,i}(x_n, a_n) \cdot b_i(z_{n+1}), \quad (2.2.6)$$

where the coordinates $\{\beta_{n,i}(x_n, a_n)\}$ are given by (2.2.5).

Instead of approximating the optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z})$ directly via nested simulation, now we are able to approximate it by estimating all its coordinates $\{\beta_{n,i}(x_n, a_n)\}$ w.r.t. $\mathcal{B}$ in (2.2.5). The difficulty lies in how to efficiently estimate the conditional expectations on the right side of (2.2.5). To avoid nested simulation, we propose the following regression scheme. The key idea is to treat $\beta_{n,i}(x_n, a_n)$ as the expected response, and thus in the regression an observation of the expected response is a sample outcome of $V_{n+1}(x_{n+1}) \cdot h_i(z_{n+1})|x_n, a_n$, which is obtained by exercising a given suboptimal policy $\widetilde{\boldsymbol{\alpha}} := (\widetilde{\alpha}_0, \widetilde{\alpha}_1, ..., \widetilde{\alpha}_{N-1})$ and computing the value function along one sample path. Further let

$$\boldsymbol{\phi}_{n,i}(x_n, a_n) \stackrel{\triangle}{=} (\phi^1_{n,i}(x_n, a_n), ..., \phi^K_{n,i}(x_n, a_n))^T$$

denote the vector of regressors (dependent variables) in the regression, and later we will illustrate on how to properly construct them based on the choice of $\mathcal{B}$. The linear regression model could be formulated as

$$V_{n+1}(x_{n+1}) \cdot h_i(z_{n+1})|x_n, a_n = (\boldsymbol{\phi}_{n,i}(x_n, a_n))^T \boldsymbol{\theta}_{n,i} + \epsilon_{n,i}, \quad (2.2.7)$$

where $\boldsymbol{\theta}_{n,i} := (\theta^1_{n,i}, ..., \theta^K_{n,i})^T$ is the vector of regression coefficients and $\epsilon_{n,i}$ is the noise of the regression model. The complete algorithm is summarized as follows, which is referred to as Algorithm "ODPAR".

For convenience, we refer to $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$ as the *regression-based (dual) penalty*. Note that one could implement Algorithm 2.2.1 efficiently by reusing the sample paths generated to evaluate the suboptimal policy in the primal problem. Hence, minimal extra simulation or computational costs are required. Moreover, generating a regression-based dual penalty $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$ as in (2.2.9) does not incur nested simulation due to the relief of conditional expectations. Hence, computational efficiency is significantly improved.

To have an implementable algorithm, the functional basis $\mathcal{B}$ should possess the following properties: completeness, orthogonality and countability. The reason is that completeness guarantees (2.2.3) to hold, orthogonality significantly simplifies the parameters of linear system (2.2.4), and countability determines whether the linear system is solvable. Moreover, one would prefer a basis $\mathcal{B}$ such that the resulted regression-based dual penalty $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$ is feasible, and thus induces valid upper bounds.

Following these guidelines, we consider two natural choices of $\mathcal{B}$ as follows. The first one is the orthonormal basis of the Hilbert L^2 space induced by the probability measure ρ, and we will show that it possesses good properties such as completeness, orthogonality, and countability. The second one is the basis consisting of all the centralized moments of the noise distribution, and we will show that it results in a simple linear system of equations for the coordinates.

2.2.2 Regression Approach with L^2 Orthonormal Basis

To ease the presentation, let us first lay out some preliminaries for L^2 space. Recall the random noises $\{z_n\}$ are i.i.d. with probability measure ρ on support space $\Xi \subseteq \mathbb{R}^d$.

Algorithm 2.2.1 **O**ptimal **D**ual **P**enalty **A**pproximation via **R**egression

Input: Functional basis $\mathcal{B}$ and suboptimal policy $\widetilde{\boldsymbol{\alpha}}$.

Output: $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$—Approximation of the optimal dual penalty. 1. **Initialization**: Simulate M independent (state-action) sample paths under the given policy $\widetilde{\boldsymbol{\alpha}}$ subject to state dynamics (2.1.1). Denote the sample paths by

$$\{(x_0^j, a_0^j; x_1^j, a_1^j; ..., x_{N-1}^j, a_{N-1}^j; x_N^j) : j = 1, ..., M\},$$

where the noise sequences are $\{(z_1^j, z_2^j, ..., z_N^j) : j = 1, ..., M\}$ and $a_n^j = \widetilde{\alpha}_n(x_n^j)$. Calculate the corresponding value functions along each sample path, denoted by

$$\{(V_0^j(x_0^j), V_1^j(x_1^j), ..., V_N^j(x_N^j)) : j = 1, ..., M\},$$

where

$$V_n^j(x_n^j) \triangleq \sum_{k=n}^{N-1} r_n(x_n^j, a_n^j) + r_N(x_N^j).$$

2. **Iteration**: For $n = 0, ..., N-1$, $i \in I$, estimate the coordinate $\beta_{n,i}(x_n, a_n)$ by regression. In the regression the responses are

$$\{V_{n+1}^j(x_{n+1}^j) \cdot h_i(z_{n+1}^j) : j = 1, ..., M\}$$

and the design matrix is

$$(\boldsymbol{\phi}_{n,i}^T(x_n^1, a_n^1), ..., \boldsymbol{\phi}_{n,i}^T(x_n^M, a_n^M))^T,$$

where

$$\boldsymbol{\phi}_{n,i}(x_n^j, a_n^j) = (\phi_{n,i}^1(x_n^j, a_n^j), ..., \phi_{n,i}^K(x_n^j, a_n^j))^T.$$

Let $\widetilde{\boldsymbol{\theta}}_{n,i} := (\widetilde{\theta}_{n,i}^1, ..., \widetilde{\theta}_{n,i}^K)^T$ denote the vector of regression coefficients, i.e.,

$$\widetilde{\boldsymbol{\theta}}_{n,i} = \arg\min_{\boldsymbol{\theta}_{n,i}} \frac{1}{M} \sum_{j=1}^{M} \left(V_{n+1}^j(x_{n+1}^j) \cdot h_i(z_{n+1}^j) - \left(\boldsymbol{\phi}_{n,i}(x_n^j, a_n^j)\right)^T \boldsymbol{\theta}_{n,i} \right)^2.$$

3. **Termination**: Let

$$\widetilde{\beta}_{n,i}(x_n, a_n) \triangleq (\boldsymbol{\phi}_{n,i}(x_n, a_n))^T \widetilde{\boldsymbol{\theta}}_{n,i}. \tag{2.2.8}$$

be the estimation of $\beta_{n,i}(x_n, a_n)$ and

$$\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z}) \triangleq \sum_{n=0}^{N-1} \sum_{i \in I} \widetilde{\beta}_{n,i}(x_n, a_n) \cdot b_i(z_{n+1}), \text{ where } a_n = \alpha_n(x_n) \tag{2.2.9}$$

be the approximation of $M^*(\boldsymbol{\alpha}, \mathbf{z})$.

Consider the L^2 space induced by the measure ρ, denoted by $L^2(\rho)$. In particular, $L^2(\rho)$ is the space consisting of all ρ-measurable scalar function f such that $|f|^2$ is integrable. That is, $L^2(\rho) = \{f : \int_\Xi |f|^2 d\rho < \infty\}$.

By Theorem 4.1.3 in [16], $L^2(\rho)$ is a Banach (hence complete) space equipped with the L^2-norm $\|\cdot\|_2$, where

$$\|f\|_2 = \sqrt{\int_\Xi |f|^2 d\rho}, \quad \forall f \in L^2(\rho).$$

Here note that a metric space S is called complete if every Cauchy sequence of points in S has a limit in S. A nice property of $L^2(\rho)$ is that the L^2 norm $\|\cdot\|_2$ is generated by the inner product defined by

$$\langle f, g \rangle \triangleq \int_\Xi f \cdot g d\rho, \quad \forall f, g \in L^2(\rho).$$

Obviously, $\|f\|_2 = \sqrt{\langle f, f \rangle}$. Therefore, $L^2(\rho)$ is a Hilbert space (see, e.g., page 255 in [16]) equipped with the inner product defined above. Now let us introduce the definition of *orthonormal basis* in a Hilbert space.

Definition 2.2.1. *(***Orthonormal Basis***) For a Hilbert space H equipped with inner product $\langle \cdot, \cdot \rangle_H$ and norm $\|\cdot\|_H$, a family of mutually orthogonal unit vectors $\mathcal{E} = \{e_i \in H : i \in I\}$ is called an (complete) orthonormal basis of H if*

(i) $\|e_i\|_H = 1, \ \forall i \in I; \ \langle e_i, e_j \rangle_H = 0, \ \forall i, j \in I, i \neq j.$

(ii) Every element $f \in H$ can be represented as the linear combination of $\{e_i\}$, i.e.,

$$f \overset{H}{=} \sum_{i \in I} \beta_i e_i, \ \text{or equivalently,} \ \left\| f - \sum_{i \in I} \beta_i e_i \right\|_H = 0,$$

where at most countably many coefficients β_i may be nonzero.

It immediately follows that if $f \overset{H}{=} \sum_i \beta_i e_i$, then

$$\beta_i = \langle f, e_i \rangle_H \quad \text{and} \quad \|f\|_H = \sqrt{\sum_i |\beta_i|^2}.$$

For a general Hilbert space H, by Zorns lemma (see, e.g., [16]), there always exists an orthonormal basis; however, the number of basis functions in that orthonormal basis might be uncountable. Fortunately, for the specific Hilbert space $L^2(\rho)$, we have the following proposition on the countability of its orthonormal basis.

Proposition 2.2.1. *Let $\mathcal{E} = \{e_i \in L^2(\rho) : i \in I\}$ be an orthonormal basis of $L^2(\rho)$, then $\mathcal{E}$ is countable.*

Proof. Note that a metric space S is called separable if it contains a countable and everywhere dense subset. By Corollary 4.2.2 in [16], $L^2(\rho)$ is a separable Hilbert space. Therefore, by Corollary 4.3.4 in [16], any orthonormal basis of $L^2(\rho)$ is countable. □

Proposition 2.2.1 implies that we can rewrite the orthonormal basis $\mathcal{E}$ as $\mathcal{E} = \{e_i \in L^2(\rho) : i = 0, 1, 2, ..., \}$. Without loss of generality, let us assume $e_0 \equiv 1$; otherwise we could simply perform the Gram–Schmidt process to achieve that. Note that the basis $\mathcal{E}$ is uniquely determined when $e_0 \equiv 1$. To gain more intuition, below are examples of orthonormal basis of $L^2(\rho)$ for several common probability measures.

Example 1. Finite Discrete Distribution. Assume ρ is a finite discrete probability measure with positive probability masses on $\Xi = \{y_0, ..., y_{p-1}\}$, a finite discrete set of p points. It is easy to verify that the cardinality of any orthonormal basis of $L^2(\rho)$ is p. In fact,

$$\mathcal{G} = \{g_i : g_i(y) = \frac{1}{\sqrt{\rho(y_i)}} \mathbb{1}\{y = y_i\},\ i = 0, 1, ..., p - 1.\}$$

is an orthonormal basis of $L^2(\rho)$.

Example 2. Standard Normal Distribution. Assume ρ is the measure induced by a standard normal distribution such that $\rho(z) = 1/\sqrt{2\pi}\exp(-\frac{z^2}{2}), \forall z \in \Xi = \mathbb{R}$, then the Hermite polynomials

$$\{1;\ z;\ \frac{1}{\sqrt{2}}\left(z^2 - 1\right);\ \frac{1}{\sqrt{6}}\left(z^3 - 3z\right); ...\}$$

is a countable orthonormal basis of $L^2(\rho)$.

Example 3. Standard Exponential Distribution. Assume ρ is the measure induced by a standard exponential distribution such that $\rho(z) = \exp(-z), \forall z \in \Xi = \mathbb{R}^+$, then the Laguerre polynomials

$$\{1;\ z-1;\ \frac{1}{2}\left(z^2-4z+2\right);\ \frac{1}{6}\left(-z^3+9z^2-18z+6\right);...\}$$

is a countable orthonormal basis of $L^2(\rho)$.

Next, we will show that the L^2 orthonormal basis $\mathcal{E}_1 := \mathcal{E} \setminus \{e_0\}$ is a valid choice of basis $\mathcal{B}$ in the proposed regression approach. In view of the state dynamics (2.1.1), for a fixed (x_n, a_n) pair, $V_{n+1}(x_{n+1}) = V_{n+1}(f(x_n, a_n, z_{n+1}))$ is a random variable as a function of the random variable z_{n+1}. Further suppose $V_{n+1}(x_{n+1})$ has finite second moment, i.e., $V_{n+1}(f(x_n, a_n, \cdot)) \in L^2(\rho)$. Since $\mathcal{E} = \{e_i : i = 0, 1, ...\}$ with $e_0 \equiv 1$ is an orthonormal basis of $L^2(\rho)$, we can express $V_{n+1}(x_{n+1})$ w.r.t. $\mathcal{E}$, i.e.,

$$V_{n+1}(x_{n+1}) = \sum_{i=0}^{\infty} \delta_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}), \tag{2.2.10}$$

where, following Definition 2.2.1, $\{\delta_{n,i}(x_n, a_n)\}$ are coordinates such that

$$\begin{aligned} \delta_{n,i}(x_n, a_n) &= \langle V_{n+1}(x_{n+1}), e_i(z_{n+1})\rangle = \int_{\Xi} V_{n+1}(x_{n+1}) \cdot e_i(z)\rho(dz) \\ &= \mathbb{E}\left[V_{n+1}(x_{n+1}) \cdot e_i(z_{n+1})|x_n, a_n\right],\ i = 0, 1, ... \end{aligned} \tag{2.2.11}$$

In particular, noting that $e_0 \equiv 1$, then we have

$$\delta_{n,0}(x_n, a_n) = \mathbb{E}\left[V_{n+1}(x_{n+1})|x_n, a_n\right]. \tag{2.2.12}$$

Combining (2.2.2), (2.2.10) and (2.2.12), we have

$$\begin{aligned} M_n^*(\boldsymbol{\alpha}, \mathbf{z}) &= \left(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1})|x_n, a_n\right]\right) \\ &= \sum_{i=0}^{\infty} \delta_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}) - \delta_{n,0}(x_n, a_n) \cdot e_0(z_{n+1}) \\ &= \sum_{i=1}^{\infty} \delta_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}). \end{aligned} \tag{2.2.13}$$

Thus, the optimal dual penalty could be rewritten as

$$M^*(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{n=0}^{N-1} \sum_{i=1}^{\infty} \delta_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}), \text{ where } a_n = \alpha_n(x_n), \tag{2.2.14}$$

where $\delta_{n,i}(x_n, a_n)$ is given by (2.2.11). Comparing (2.2.6) with (2.2.14), we can see that the L^2 orthonormal basis $\mathcal{E}_1 = \{e_1, e_2, ...\}$ is a valid choice of $\mathcal{B}$ in the regression approach. Moreover, the function $h_i(\cdot)$ in (2.2.5) simply reduces to $e_i(\cdot)$ due to $\mathcal{E}_1$'s orthogonality. Hence, the corresponding coordinates $\{\delta_{n,i}(x_n, a_n)\}$ are straightforward to compute as in (2.2.11). Next let us show that using $\mathcal{E}_1$ in the regression approach induces feasible dual penalties, and thus valid upper bounds on the optimal value function. To be more specific, let

$$\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z}) \stackrel{\triangle}{=} \sum_{n=0}^{N-1} \sum_{i=1}^{P} \widetilde{\delta}_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}), \text{ where } a_n = \alpha_n(x_n) \tag{2.2.15}$$

denote the regression-based dual penalty w.r.t. the basis $\mathcal{E}_1$, where P is the order of the basis truncation and $\{\widetilde{\delta}_{n,i}(x_n, a_n)\}$ is the estimation of $\{\delta_{n,i}(x_n, a_n)\}$ after regression. The following theorem shows the dual feasibility of $\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z})$.

Theorem 2.2.1. *The regression-based dual penalty $\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z})$ defined in (2.2.15) is a feasible dual penalty, i.e., $\mathbb{E}_0\left[\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z})\right] = 0, \ \forall \boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}$.*

Proof. Notice that $\forall n = 0, ..., N-1, i = 1, 2, ...$

$$\mathbb{E}\left[e_i(z_{n+1})\right] = \int_{\Xi} e_i(z)\rho(dz) = \int_{\Xi} e_i(z) \cdot e_0(z)\rho(dz) = \langle e_i(z_{n+1}), e_0(z_{n+1})\rangle = 0,$$

where we use the fact $e_0(z_{n+1}) \equiv 1$ and $\mathcal{E}$'s orthogonality. Then we have $\forall \boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}$

$$\begin{aligned}
&\mathbb{E}_0\left[\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z})\right] = \mathbb{E}\left[\widetilde{M}^l(\boldsymbol{\alpha}, \mathbf{z})\middle|\mathscr{F}_0\right] = \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{i=1}^{P} \widetilde{\delta}_{n,i}(x_n, a_n) \cdot e_i(z_{n+1})\middle|\mathscr{F}_0\right] \\
\stackrel{(i)}{=}\ &\mathbb{E}\Bigg[\mathbb{E}\Bigg[\sum_{n=0}^{N-2}\sum_{i=1}^{P} \widetilde{\delta}_{n,i}(x_n, a_n) \cdot e_i(z_{n+1}) \\
&+ \sum_{i=1}^{P} \widetilde{\delta}_{N-1,i}(x_{N-1}, a_{N-1}) \cdot e_i(z_N)\Big|\mathscr{F}_{N-1}\Bigg]\Big|\mathscr{F}_0\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{n=0}^{N-2}\sum_{i=1}^{P}\widetilde{\delta}_{n,i}(x_n,a_n)\cdot e_i(z_{n+1})\Big|\mathscr{F}_{N-1}\right]\Big|\mathscr{F}_0\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{P}\widetilde{\delta}_{N-1,i}(x_{N-1},a_{N-1})\cdot e_i(z_N)\Big|\mathscr{F}_{N-1}\right]\Big|\mathscr{F}_0\right] \\
&\overset{(ii)}{=} \mathbb{E}\left[\sum_{n=0}^{N-2}\sum_{i=1}^{P}\widetilde{\delta}_{n,i}(x_n,a_n)\cdot e_i(z_{n+1})\Big|\mathscr{F}_0\right] \\
&\quad +\mathbb{E}\left[\sum_{i=1}^{P}\widetilde{\delta}_{N-1,i}(x_{N-1},a_{N-1})\cdot \mathbb{E}\left[e_i(z_N)\big|\mathscr{F}_{N-1}\right]\Big|\mathscr{F}_0\right] \\
&\overset{(iii)}{=} \mathbb{E}\left[\sum_{n=0}^{N-2}\sum_{i=1}^{P}\widetilde{\delta}_{n,i}(x_n,a_n)\cdot e_i(z_{n+1})\Big|\mathscr{F}_0\right] \\
&\cdots \\
&= \mathbb{E}\left[\sum_{i=1}^{P}\widetilde{\delta}_{0,i}(x_0,a_0)\cdot e_i(z_1)\Big|\mathscr{F}_0\right] = \sum_{i=1}^{P}\widetilde{\delta}_{0,i}(x_0,a_0)\cdot \mathbb{E}\left[e_i(z_1)\big|\mathscr{F}_0\right] \\
&= 0,
\end{aligned}
$$

where (i) follows from the tower property of conditional expectations and $\mathscr{F}_0 \subseteq \mathscr{F}_{N-1}$, (ii) follows from the fact that $\boldsymbol{\alpha}$ is $\mathcal{F}$-adapted, and thus $\{\widetilde{\delta}_{N-1,i}(x_{N-1},a_{N-1})\}$ are $\mathscr{F}_{N-1}$-measurable, and finally (iii) follows from the fact that z_N is independent of $\mathscr{F}_{N-1}$, and thus $\mathbb{E}\left[e_i(z_N)|\mathscr{F}_{N-1}\right] = \mathbb{E}\left[e_i(z_N)\right] = 0$. □

From the proof, we can see the fact that each basis function $e_i(\cdot)$ has expectation zero w.r.t. the probability measure ρ is essential for $\widetilde{M}^l(\boldsymbol{\alpha},\mathbf{z})$ to maintain the dual feasibility. Theorem 2.2.1 implies that the dual problem (2.1.5) with penalty $\widetilde{M}^l(\boldsymbol{\alpha},\mathbf{z})$ provides a valid upper bound on the optimal value function, i.e., $V_0^{\widetilde{M}^l}(x_0) \geq V_0(x_0), \forall x_0 \in \mathscr{X}_0$. The tightness of the upper bound $V_0^{\widetilde{M}}(x_0)$ is directly affected by the accuracy of the penalty approximation. If zero regression error is assumed, then $\widetilde{M}^l(\boldsymbol{\alpha},\mathbf{z})$ converges to $M(\boldsymbol{\alpha},\mathbf{z})$ in L^2 as the order of basis truncation, P, goes to infinity, due to the completeness of $\mathcal{E}$. When regression error is taken into account, the construction of regressors $\boldsymbol{\phi}$ is very important in controlling that error. We will discuss this issue later.

Although, in theory, the unique L^2 orthonormal basis $\mathcal{E}_1$ could be calculated sequentially via Gram-Schmidt process for any probability measure ρ as long as it induces a Hilbert space $L^2(\rho)$. In practice, this procedure might be complicated for complex probability measures, which is the drawback of using L^2 orthonormal basis in the regression approach. Next, we will derive an alternative functional basis that is easy to calculate. It is induced by carrying out Taylor series expansion on the value function, and consisting of centralized moments of the noise distribution along each dimension.

2.2.3 Regression Approach with Taylor Series Basis

Now let us derive the functional basis of $\mathbb{M}_{\mathscr{F}}$ induced by Taylor series expansion on the value function. To ease the presentation, let us further assume the state dynamics is linear in noise as follows.

$$x_{n+1} = b(x_n, a_n) + \sigma(x_n, a_n) z_{n+1}, n = 0, 1, ..., N-1, \tag{2.2.16}$$

where $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are deterministic $\mathscr{F}$-measurable functions, and z_{n+1} has finite moments (up to some order). Although the derivation could be easily extended to the setting with general state dynamics (2.1.1), we confine it within (2.2.16) to better demonstrate that several existing approaches could be regarded as special cases of the proposed approach. Note that the state dynamics (2.2.16) covers a wide range of dynamic programming problems. For example, it has been widely used when modeling dynamics of asset/stock prices in finance literature, where, for instance, x might represent the asset prices, $b(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ might represent the drifts and the volatilities in the asset prices, respectively. The reward functions and the objective remain the same. Therefore, the primal problem is solved as previous. We will focus on deriving a functional basis of the dual penalty space that facilitates the approximation of optimal dual penalty.

For convenience, let $\widehat{x}_{n+1} := b(x_n, a_n)$, i.e., $\widehat{x}_{n+1}$ represents the expected future

state at period $(n+1)$ given state x_n and action a_n at period n. In particular, $\widehat{x}_{n+1}$ is $\mathscr{F}_n$-measurable. Rewrite the state dynamics (2.2.16) as

$$x_{n+1} - \widehat{x}_{n+1} = \sigma(x_n, a_n) z_{n+1},\ n = 0, 1, ..., N-1.$$

Consider the Taylor series expansion on $V_{n+1}(x_{n+1})$ around the point $\widehat{x}_{n+1}$ up to R-th order

$$V_{n+1}(x_{n+1}) \approx V_{n+1}(\widehat{x}_{n+1}) + \sum_{r=1}^{R} \frac{1}{r!} \frac{\partial^r}{\partial x^r} V_{n+1}(\widehat{x}_{n+1}) \cdot (\sigma)^r (z_{n+1})^r, \tag{2.2.17}$$

where we use σ as an abbreviation of $\sigma(x_n, a_n)$. Taking expectations w.r.t. $\mathscr{F}_n$ on both sides of (2.2.17), we have

$$\mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n\right] \approx V_{n+1}(\widehat{x}_{n+1}) + \sum_{r=1}^{R} \frac{1}{r!} \frac{\partial^r}{\partial x^r} V_{n+1}(\widehat{x}_{n+1}) \cdot (\sigma)^r \mathbb{E}\left[(z_{n+1})^r\right], \tag{2.2.18}$$

where we use the fact that $V_{n+1}(\widehat{x}_{n+1})$ and its partial derivatives, and $\sigma(x_n, a_n)$ are $\mathscr{F}_n$-measurable. Subtracting (2.2.18) from (2.2.17), we have

$$V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n\right] \approx \sum_{r=1}^{R} \frac{1}{r!} \frac{\partial^r}{\partial x^r} V_{n+1}(\widehat{x}_{n+1}) \cdot (\sigma)^r \Big((z_{n+1})^r - \mathbb{E}\big[(z_{n+1})^r\big] \Big).$$

The above approximation has a nice structure because the expectations are on $(z_{n+1})^r$ instead of $V_{n+1}(x_{n+1})$, which could be calculated analytically. However, the partial derivatives could be difficult to compute directly and often require approximation methods such as finite difference. This inspires us to consider the functional basis

$$\mathcal{D} \stackrel{\triangle}{=} \{d_r : d_r(z) = (z^r - \mathbb{E}[z^r]),\ r = 1, ..., R\}$$

of the dual penalty space. It follows that the optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z})$ could be expressed w.r.t. $\mathcal{D}$ as

$$\begin{aligned} M^*(\boldsymbol{\alpha}, \mathbf{z}) &\approx \sum_{n=0}^{N-1} \sum_{r=1}^{R} \gamma_{n,r}(x_n, a_n) \cdot \Big((z_{n+1})^r - \mathbb{E}\big[(z_{n+1})^r\big] \Big) \\ &= \sum_{n=0}^{N-1} \sum_{r=1}^{R} \gamma_{n,r}(x_n, a_n) \cdot d_r(z_{n+1}), \end{aligned}$$

where $\{\gamma_{n,r}(x_n, a_n)\}$ are the corresponding coordinates. Following the derivations in (2.2.3)-(2.2.4), we obtain a simple linear system of equations for $\{\gamma_{n,r}(x_n, a_n)\}$:

$$\begin{bmatrix} \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big(z_{n+1} - \mathbb{E}[z_{n+1}]\big)\big] \\ \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big((z_{n+1})^2 - \mathbb{E}\left[(z_{n+1})^2\right]\big)\big] \\ \vdots \\ \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big((z_{n+1})^R - \mathbb{E}\left[(z_{n+1})^R\right]\big)\big] \end{bmatrix} \approx \left[Cov \begin{pmatrix} z_{n+1} \\ (z_{n+1})^2 \\ \vdots \\ (z_{n+1})^R \end{pmatrix} \right] \begin{bmatrix} \gamma_{n,1}(x_n, a_n) \\ \gamma_{n,2}(x_n, a_n) \\ \vdots \\ \gamma_{n,R}(x_n, a_n) \end{bmatrix}, \tag{2.2.19}$$

where $Cov(\cdot)$ represents the covariance matrix of a random vector. Note that it is nonsingular, and thus we can solve (2.2.19) as

$$\begin{bmatrix} \gamma_{n,1}(x_n, a_n) \\ \gamma_{n,2}(x_n, a_n) \\ \vdots \\ \gamma_{n,R}(x_n, a_n) \end{bmatrix} \approx \left[Cov \begin{pmatrix} z_{n+1} \\ (z_{n+1})^2 \\ \vdots \\ (z_{n+1})^R \end{pmatrix} \right]^{-1} \begin{bmatrix} \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big(z_{n+1} - \mathbb{E}[z_{n+1}]\big)\big] \\ \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big((z_{n+1})^2 - \mathbb{E}\left[(z_{n+1})^2\right]\big)\big] \\ \vdots \\ \mathbb{E}_n\big[V_{n+1}(x_{n+1})\big((z_{n+1})^R - \mathbb{E}\left[(z_{n+1})^R\right]\big)\big] \end{bmatrix}. \tag{2.2.20}$$

For simplicity, let us rewrite (2.2.20) explicitly as

$$\gamma_{n,r}(x_n, a_n) \approx \mathbb{E}\left[V_{n+1}(x_{n+1}) l_r(z_{n+1}) | x_n, a_n\right], \ r = 1, ..., R,$$

where it is esay to see that $\{l_r(\cdot)\}$ are simple polynomial functions of degrees $\leq R$. Therefore, the coordinates $\{\gamma_{n,r}(x_n, a_n)\}$ could be easily computed, which is the main advantage of using the function basis $\mathcal{D}$ in the regression approach. The drawback is that it does not guarantee convergence of the optimal dual penalty approximation as R goes to infinity, since in general Taylor series expansion does not converge. Next, similar to using the function basis $\mathcal{E}_1$, let us show that using $\mathcal{D}$ in the regression approach induces feasible dual penalties, and thus valid upper bounds on the optimal value function. To this end, let

$$\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}) \triangleq \sum_{n=0}^{N-1} \sum_{r=1}^{R} \widetilde{\gamma}_{n,i}(x_n, a_n) \cdot d_i(z_{n+1}), \text{ where } a_n = \alpha_n(x_n) \tag{2.2.21}$$

denote the regression-based dual penalty generated via Algorithm 2.2.1 with $\mathcal{D}$ as the functional basis $\mathcal{B}$, where $\{\widetilde{\gamma}_{n,i}(x_n, a_n)\}$ is the estimation of $\{\gamma_{n,i}(x_n, a_n)\}$ through regression. The following theorem shows the dual feasibility of $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z})$.

Theorem 2.2.2. *The regression-based dual penalty* $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z})$ *defined in (2.2.21) is a feasible dual penalty, i.e.,* $\mathbb{E}_0\left[\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z})\right] = 0,\ \forall \boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}$.

Noting that each basis function d_i in $\mathcal{D}$ has expectation zero w.r.t. the probability measure ρ, the proof is similar to the proof of Theorem 2.2.1. Similarly, an immediate implication of Theorem 2.2.2 is that the dual problem (2.1.5) with penalty $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z})$ provides a valid upper bound on the optimal value function, i.e., $V_0^{\widetilde{M}^t}(x_0) \geq V_0(x_0), \forall x_0 \in \mathscr{X}_0$. Again, the tightness of the upper bound $V_0^{\widetilde{M}}(x_0)$ is directly affected by the accuracy of the approximation $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z})$, which heavily relies on the regressors $\boldsymbol{\phi}$ constructed in the regression. We will discuss this issue next.

Remark 2.2.1. *Interestingly, one could verify that if the noises* $\{z_n\}$ *are symmetric, then the first two basis functions* $d_1, d_2 \in \mathcal{D}$ *coincide with the first two basis functions* $e_1, e_2 \in \mathcal{E}_1$ *(after normalization), respectively. For example, when* $\{z_n\}$ *follows a standard normal distribution, then* $\{d_1 = z, d_2 = (z^2 - 1)\} \subset \mathcal{D}$ *coincides with* $\{e_1 = z, e_2 = \frac{1}{\sqrt{2}}(z^2 - 1)\} \subset \mathcal{E}_1$ *in Example 2 of Chapter 2.2.2.*

2.2.4 Implementation

We will discuss two important implementation issues regarding the proposed regression approach, i.e., how to properly construct the regressors in the regression and how to solve the inner optimization problem in the resulted dual problem.

As mentioned previously, the accuracy of the regression-based dual penalty $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$ heavily relies on the choice of regressors $\boldsymbol{\phi}$ for regressing the coordinates β. The general guideline is to construct $\boldsymbol{\phi}$ similar to β in structure. In view of the relationship

$$\beta_{n,i}(x_n, a_n) = \mathbb{E}\left[V_{n+1}(x_{n+1}) \cdot h_i(z_{n+1}) | x_n, a_n\right],$$

a natural candidate regressor is $\mathbb{E}\left[\widetilde{V}_{n+1}(x_{n+1}) \cdot h_i(z_{n+1}) | x_n, a_n\right]$, where $\widetilde{V}_{n+1}(\cdot)$ is an approximation of $V_{n+1}(\cdot)$ with a closed-form expression. For example, $\widetilde{V}_{n+1}(\cdot)$ could be the value function induced by a naive policy, or an approximate optimal value

function, or the value function of a simpler DP (e.g., through a relaxation of the constraints on the actions of the original DP). Multiple regressors of such form could be included in the regression.

More specifically, if the Taylor series basis $\mathcal{D}$ is used, then there exists good regressors of a simpler form. Note that the coordinates $\{\gamma_{n,r}(x_n, a_n)\}$ take the positions of the partial derivatives in the Taylor series basis expansion of the optimal dual penalty. Hence, a good candidate regressor is $\frac{\partial^r}{\partial x^r}\widetilde{V}_{n+1}(\widehat{x}_{n+1}) \cdot (\sigma)^r$, where $\widetilde{V}_{n+1}(\cdot)$ admits partial derivatives that are easy to derive. Again, $\widetilde{V}_{n+1}(\cdot)$ could be the value function induced by a naive policy, or an approximate optimal value function, or the value function of a simpler DP.

Theoretically, after plugging the regression-based dual penalty $\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z})$ into the dual problem (2.1.5), we only need to solve a series of deterministic inner optimization problem to generate an upper bound estimator. However, they might be hard to solve without good structural properties such as convexity. Therefore, for a convex DP, ideally the dual penalty plugged in the dual problem should preserve convexity in the inner optimization problem. Note that the optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z})$ might not be convex even if the value functions $\{V_n(x_n)\}$ are concave, and in this case plugging in the optimal dual penalty causes the inner optimization problem to lose convexity. One solution to this issue is to linearize the optimal dual penalty around a fixed policy. The resulted optimal dual penalty approximation is affine in policy, and thus preserves convexity in the dual problem.

The proposed framework of regression approach does not suffer from this issue. It is very robust in the sense that an optimal dual penalty approximation with desired structural properties could be generated by using proper regressors, noting that the resulted dual penalty is a linear combination of the regressors. For example, for a convex DP, the regression-based dual penalty preserves convexity in the dual problem by constructing regressors that are affine in policy.

2.3 Special Cases

It turns out, in a broader sense, several existing approaches to approximating the optimal dual penalty are special cases of the proposed regression approach under specific settings.

2.3.1 American Option Pricing

American-style option pricing problem is one of the simplest non-trivial DPs. Various methodologies and numerical methods are first proposed under the setting of option pricing, and then generalized to general DPs. We will demonstrate that the non-nested simulation approach developed by [12] and [89] to approximating the optimal dual martingale in American option pricing could be regarded as a special case of the proposed regression approach with the first-order Taylor series basis $\{d_1\}$.

Let us first briefly describe the American option pricing problem. Assume the asset price $X(t)$ satisfies a stochastic differential equation (SDE) w.r.t. Brownian motion as follows.

$$dX(t) = b(t, X(t))\, dt + \sigma(t, X(t))\, dW(t),$$

where $t \in [0, T]$, $X(t)$ denotes the asset price at time t with given initial deterministic value $X(0) = X_0$, $W(t)$ represents the (standard) Brownian motion, and the coefficients b, σ are functionals satisfying mild regularity conditions. By convention, we use $\{\mathscr{F}_t : 0 \leq t \leq T\}$ to denote the augmented information filtration generated by the Brownian motion $W(t)$. Consider a Bermudan option on $X(t)$ that can be exercised at any date from the time set $\mathcal{T} = \{T_0, T_1, ..., T_N\}$, with $0 = T_0 < T_1 < \cdots < T_N = T$. When exercising at time $T_n \in \mathcal{T}$, the option holder receives a payoff $H_{T_n} = h(T_n, X(T_n))$. The goal is to evaluate the price of the Bermudan option, that is, to find

$$Primal: \qquad V_0 = \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[h(\tau, X(\tau)) \,|X(0) = X_0\right],$$

where τ is an exercising strategy, i.e., a stopping time adapted to the filtration $\mathcal{F} = \{\mathscr{F}_{T_n} : n = 0, ..., N\}$ taking values in $\mathcal{T}$, and V_0 denotes the Bermudan option price at time T_0 with initial asset price X_0.

As for the dual formulation, [3] and [44] show that all the $\mathcal{F}$-adapted martingales in $\mathbb{M}_{\mathcal{F}} = \{M = (M_0, ..., M_N) : M \text{ is } \mathcal{F}\text{-adapted}\}$ are feasible dual penalties. In particular, the optimal dual martingale M^* is the Doob-Meyer martingale component of the Bermudan price process $\{V_{T_n} : n = 0, ..., N\}$, i.e.,

$$\begin{cases} M_0^* = 0, \\ M_{T_{n+1}}^* = M_{T_n}^* + V_{T_{n+1}} - \mathbb{E}_{T_n}\left[V_{T_{n+1}}\right], n = 0, ..., N-1. \end{cases} \tag{2.3.1}$$

The question is how to efficiently estimate M^* in a non-nested manner. Note that the martingale M^* is driven by Brownian motion, by martingale representation theorem we have

$$M_{T_n}^* = \int_0^{T_n} C_s dW(s), \qquad n = 0, ..., N,$$

where C_s is a predictable process w.r.t. the filtration $\mathcal{F}$. [12] proposes an Ito sum approximation scheme as follows.

$$M_{T_n}^* = \int_0^{T_n} C_s dW(s) \approx \sum_{i=0}^{n-1} C_{T_i}\left(\Delta W_{T_{i+1}}\right), n = 0, ..., N,$$

where $\left(\Delta W_{T_{i+1}}\right) := (W(T_{i+1}) - W(T_i))$ denotes the Brownian increment. Combining with (2.3.1), we have

$$V_{T_{n+1}} - \mathbb{E}_{T_n}\left[V_{T_{n+1}}\right] = M_{T_{n+1}}^* - M_{T_n}^* \approx C_{T_n}\left(\Delta W_{T_{n+1}}\right), \ n = 0, ..., N-1. \tag{2.3.2}$$

Multiplying both sides of (2.3.2) by the Brownian increment $\left(\Delta W_{T_{n+1}}\right)$ and taking conditional expectations w.r.t. $\mathscr{F}_{T_n}$, we obtain

$$C_{T_n} \approx \frac{1}{T_{n+1} - T_n}\mathbb{E}_{T_n}\left[V_{T_{n+1}}\left(\Delta W_{T_{n+1}}\right)\right], \ n = 0, ..., N-1.$$

Starting from here, regression is further applied to estimate C_{T_n} in order to avoid nested simulation. A regression-based dual martingale $\widetilde{M} = (\widetilde{M}_{T_0}, \widetilde{M}_{T_1}, ..., \widetilde{M}_{T_N})$ is

given by

$$\widetilde{M}_{T_n} \triangleq \sum_{i=0}^{n-1} \widetilde{C}_{T_i} \cdot \left(\Delta W_{T_{i+1}}\right), \quad n = 0, ..., N, \tag{2.3.3}$$

where $\widetilde{C}_{T_i}$ is the estimation of C_{T_i} after regression. Since an Ito sum approximation of a stochastic integral could be viewed as a first-order Taylor series expansion type of scheme, approximation scheme (2.3.3) could be viewed as a special case of the proposed regression approach with the first-order Taylor series basis $\{d_1\}$.

Remark 2.3.1. *In [12], the authors actually consider a finer Ito sum approximation of the optimal dual martingale by introducing a finer partition of the time span. Here we present a simpler version of their approach so that its connection with our proposed approach can be better understood.*

2.3.2 Controlled Markov Diffusion

[86] studies the form of optimal dual penalty under the setting of controlled Markov diffusion (CMD) and show that it is a stochastic integral. It then inspires the authors to propose an approximation scheme for the optimal dual penalty of a discrete-time DP with state dynamics (2.2.16) as follows.

$$\widehat{M}(\boldsymbol{\alpha}, \mathbf{z}) \triangleq \sum_{n=0}^{N-1} \frac{\partial}{\partial x} V_n(x_n) \cdot \sigma(x_n, a_n) \cdot z_{n+1}, \tag{2.3.4}$$

where the partial derivative $\frac{\partial}{\partial x} V_n(x_n)$ is estimated via finite difference method. This scheme is derived by mimicking the Ito sum approximation of the optimal dual penalty (which is a stochastic integral) under the setting of CMD. The connection with our proposed regression approach is evident in the sense that the approximation scheme (2.3.4) could be interpreted as a result of a first-order Taylor series basis expansion of the optimal dual penalty, except that coordinates are estimated using finite difference method instead of regression. Another difference we must point out is that $\frac{\partial}{\partial x} V_{n+1}(\hat{x}_{n+1})$ instead of $\frac{\partial}{\partial x} V_n(x_n)$ is used in our proposed regression approach. We

will show that, at least for the following linear quadratic control (LQC) problem, our approach is better in terms of approximation accuracy.

2.3.3 Linear Quadratic Control

Now let us consider the Taylor series basis expansion of the optimal dual penalty for the classic LQC problem, since LQC problem has been extensively studied and applied, see [14]. Assuming a linear state dynamics

$$x_{n+1} = A_n x_n + B_n a_n + z_{n+1}, \quad n = 0, ..., N-1,$$

with an expected quadratic cost

$$\mathbb{E}_0 \left[x_N^T Q_N x_N + \sum_{n=0}^{N-1} \left(x_n^T Q_n x_n + a_n^T R_n a_n \right) \right], \tag{2.3.5}$$

where A_n, B_n, Q_n, R_n are known matrices of appropriate dimensions, $\{z_n\}$ are proper-dimensional i.i.d. vector random variables with zero mean and finite second moment. Further assume Q_n and R_n are positive semi-definite. The objective is to select a non-anticipative policy $\boldsymbol{\alpha}^*$ that minimizes the total cost defined in (2.3.5). It turns out that the optimal control policy (e.g., see [14]) α_n^* admits a closed-form expression $\alpha_n^*(x_n) = L_n x_n$, where the gain matrix L_n is given by $L_n = -\left(B_n^T K_{n+1} B_n + R_n\right)^{-1} B_n^T K_{n+1} A_n$, and the positive semi-definite matrix K_n is given recursively by

$$\begin{cases} K_N = Q_N, \\ K_n = A_n^T \left(K_{n+1} - K_{n+1} B_n \left(B_n^T K_{n+1} B_n + R_n \right)^{-1} B_n^T K_{n+1} \right) A_n + Q_n. \end{cases}$$

The optimal value function $V_n(x_n)$ then takes the form

$$V_n(x_n) = x_n^T K_n x_n + \sum_{i=n}^{N-1} \mathbb{E}\left[z_{n+1}^T K_{n+1} z_{n+1} \right], \ n = 0, ..., N.$$

For the corresponding dual problem via information relaxation, It is easy to verify that (see, e.g., [45]) the value-based optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z})$ takes the form

$$M^*(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{n=0}^{N-1} \Big(V_{n+1}(x_{n+1}) - \mathbb{E}\left[V_{n+1}(x_{n+1}) | x_n, a_n \right] \Big)$$

$$= \sum_{n=0}^{N-1} \Big\{ 2(A_n x_n + B_n a_n)^T K_{n+1} z_{n+1} + z_{n+1}^T K_{n+1} z_{n+1} - \mathbb{E}\left[z_{n+1}^T K_{n+1} z_{n+1} \right] \Big\}.$$

Next let us approximate $M^*(\boldsymbol{\alpha}, \mathbf{z})$ using the second-order Taylor series basis $\{d_1, d_2\}$, where we apply second-order Taylor series expansion on $V_{n+1}(x_{n+1})$ around $\widehat{x}_{n+1} = A_n x_n + B_n a_n$. That is,

$$\begin{aligned}
&\widetilde{M}(\boldsymbol{\alpha}, \mathbf{z}) \\
&= \sum_{n=0}^{N-1} \left\{ \left(\frac{\partial}{\partial x} V_{n+1}(\widehat{x}_{n+1}) \right)^T z_{n+1} + z_{n+1}^T \frac{\partial^2}{\partial x^2} V_{n+1}(\widehat{x}_{n+1}) z_{n+1} - \mathbb{E}\left[z_{n+1}^T \frac{\partial^2}{\partial x^2} V_{n+1}(\widehat{x}_{n+1}) z_{n+1} \right] \right\} \\
&= \sum_{n=0}^{N-1} \Big\{ 2(A_n x_n + B_n a_n)^T K_{n+1} z_{n+1} + z_{n+1}^T K_{n+1} z_{n+1} - \mathbb{E}\left[z_{n+1}^T K_{n+1} z_{n+1} \right] \Big\} = M^*(\boldsymbol{\alpha}, \mathbf{z}).
\end{aligned}$$

Therefore, the second-order Taylor series basis expansion of the optimal dual penalty is exact, which is due to the linearity of state dynamics and quadratic structure of the optimal value function. If using the approximation scheme in (2.3.4), we have

$$\widehat{M}(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{n=0}^{N-1} \left(\frac{\partial}{\partial x} V_n(x_n) \right)^T z_{n+1} = \sum_{n=0}^{N-1} 2 x_n^T K_n z_{n+1} \neq M^*(\boldsymbol{\alpha}, \mathbf{z}).$$

2.4 Numerical Experiments: Dynamic Trading

2.4.1 Dynamic Trading with Predictable Returns and Transaction Costs

To demonstrate the effectiveness of the proposed regression approach in generating accurate approximations of optimal dual penalty and tight upper bounds, we will study the dynamic trading problem with predictable returns and transaction costs in [38]. This model has been empirically tested in real financial markets and exhibits nice structural properties. The problem formulation is as follows.

Consider an investor that attempts to trade D securities at $t \in \{0, 1, ..., T\}$. The securities' price changes (returns) are driven by K market factors, i.e.,

$$r_{t+1} = \mu_t + B f_t + z_{t+1}^{(1)}, \ t = 0, ..., T-1,$$

where r_{t+1} is the $D \times 1$ vector of security returns at time $t+1$, μ_t is the deterministic "risk-free" return, f_t is the $K \times 1$ vector of market factors that predict returns, B is the

$D \times K$ matrix of constant factor loadings, and $\{z_{t+1}^{(1)}\}$ are i.i.d. zero-mean $D \times 1$ random vectors with covariance matrix $Var(z_{t+1}^{(1)}) = \Sigma$ that represent the unpredictable noise in the return. Further assume the market return vector f_t is already known to the investor at time t (before the trading) and it follows a self-evolving state dynamics

$$\Delta f_t \stackrel{\triangle}{=} f_{t+1} - f_t = -\Phi f_t + z_{t+1}^{(2)},\ t = 0, ..., T-1, \tag{2.4.1}$$

where Φ is a $K \times K$ matrix of mean-reversion coefficients for the factors, and $\{z_{t+1}^{(2)}\}$ are i.i.d. $K \times 1$ zero-mean random vectors with covariance matrix $Var(z_{t+1}^{(2)}) = \Psi$ that represent the shock affecting the predictors. Usually, we assume $\{z_{t+1}^{(2)}\}$ are normally distributed. Assume Ψ satisfies the standard conditions such that f is stationary. Note that the state dynamics (2.4.1) could be reformulated as $f_{t+1} = (I - \Phi) f_t + z_{t+1}^{(2)}$, where I is the $K \times K$ identity matrix.

Assuming trading is costly in the sense that it impacts the market and results in a price move. In particular, [38] argues that the transaction cost associated with trading quantity $a_t = x_t - x_{t-1}$, where x_t is the share quantity held by the investor at time t (after trading) and a_t is the trading quantity at time t, is given by $TC(a_t) = \frac{1}{2} a_t^T \Lambda a_t$. Here Λ is a $D \times D$ deterministic positive semi-definite matrix measuring the level of trading cost. The intuition is that trading a_t shares will move the price by $\frac{1}{2}\Lambda a_t$ and result in a total trading cost $(\frac{1}{2}\Lambda a_t)^T a_t$, which is exactly $TC(a_t)$.

The investor's objective is to choose a (non-anticipative) trading strategy that maximizes the total expected excess return with the trading costs and risk penalties taken into account, i.e.,

$$(P): \quad V_1(x_0, f_1) = \max_{\boldsymbol{\alpha} \in \mathbb{A}_{\mathcal{F}}} \mathbb{E}_1 \left[\sum_{t=1}^{T} \left(x_t^T B f_t - \frac{1}{2} a_t^T \Lambda a_t - \frac{\gamma}{2} x_t^T \Sigma x_t \right) \right], \tag{2.4.2}$$

where $\mathbb{A}$ describes the trading constraints such as no short-sell or complete liquidation of the initial position by the end of trading horizon T, $\mathcal{F} = \{\mathscr{F}_0, ..., \mathscr{F}_T\}$ is the natural information filtration, $x_t^T B f_t$ is the expected intermediate excess return, and γ is the risk-aversion coefficient.

Notice that the primal problem (2.4.2) falls into the realm of LQC if no trading constraints are imposed (except for complete liquidation of the initial position). Therefore, both the optimal value function and the optimal trading strategy could be solved in closed form using backward dynamic programming. For example, the optimal value function is quadratic in the state (x_{t-1}, f_t) and the optimal control is an affine function of the state. In particular, for risk-neutral ($\gamma = 0$) trading, [62] shows that the optimal value function $J_{t+1}(x_t, f_{t+1})$ satisfies

$$\begin{cases} J_T(x_{T-1}, f_T) = -\frac{1}{2} x_{T-1}^T \Lambda x_{T-1}, \\ J_t(x_{t-1}, f_t) = -\frac{1}{2} x_{t-1}^T A_{xx,t} x_{t-1} + x_{t-1}^T A_{xf,t} f_t + \frac{1}{2} f_t^T A_{ff,t} f_t + A_t, \ t = T-1, ..., 1, \end{cases} \tag{2.4.3}$$

where $A_{xx,t}$, $A_{xf,t}$, $A_{ff,t}$ and A_t are coefficient matrices of proper dimensions that satisfy the following backward recursions:

$$A_{xx,T} = \Lambda, \quad A_{xf,T} = \mathbf{0}, \quad A_{ff,T} = \mathbf{0}, \quad A_T = 0;$$

$$A_{xx,t} = -\Lambda(\Lambda + A_{xx,t+1})^{-1}\Lambda + \Lambda,$$

$$A_{xf,t} = \Lambda(\Lambda + A_{xx,t+1})^{-1}(B + A_{xf,t+1}(I - \Phi)),$$

$$A_{ff,t} = (B + A_{xf,t+1}(I - \Phi))^T(\Lambda + A_{xx,t+1})^{-1}(B + A_{xf,t+1}(I - \Phi)) + (I - \Phi)^T A_{ff,t+1}(I - \Phi),$$

$$A_t = trace(\Psi A_{ff,t+1}) + A_{t+1}.$$

Furthermore, the optimal policy $\boldsymbol{\alpha}^*$ is given by

$$\alpha_t^*(x_{t-1}) = (\Lambda + A_{xx,t+1})^{-1}(\Lambda x_{t-1} + (B + A_{xf,t+1}(I - \Phi))f_t) - x_{t-1}. \tag{2.4.4}$$

Note that after imposing trading constraints, as many real-world trading problems do, the classic theories in LQC fail and in general it is difficult to find the optimal policy. Here let us consider a very natural set of trading constraints: only sell of securities is allowed, short-sell of the securities is not allowed, and the initial position must be completely liquidated by terminal horizon. Mathematically, the constraint set $\mathbb{A}$ could be formulated as

$$\mathbb{A} \stackrel{\triangle}{=} \{\mathbf{a} = (a_1, ..., a_T) : x_t = x_{t-1} + a_t, a_t \leq 0, x_t \geq 0, t = 1, ..., T; x_T = 0\}.$$

Notice that all the constraints in $\mathbb{A}$ are linear (hence convex) in actions. Thus, the

problem maintains to be tractable when deterministic convex objective functions are of interest.

In the existence of trading constraints $\mathbb{A}$, the optimal trading strategy is not readily available. Alternatively, we focus on finding sufficiently good suboptimal strategies. Although the optimal policy of the unconstrained problem is unlikely to be feasible for the constrained problem (since one of the constraints in $\mathbb{A}$ might be violated), a natural feasible policy could be derived by projecting the optimal policy of the unconstrained problem onto the constraint set $\mathbb{A}$. In particular, we consider the following projected linear quadratic control (PLQC) policy $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, ..., \widehat{\alpha}_T)$ with

$$\widehat{\alpha}_t(x_{t-1}) = \max\{-x_{t-1},\ \min\{0,\ \alpha_t^*(x_{t-1})\}\},\ t = 1, ..., T, \tag{2.4.5}$$

where the projection is componentwise and $\alpha_t^*(x_{t-1})$ is given by (2.4.4). From (2.4.5), the projection of control onto $\mathbb{A}$ is easy to implement, and thus it is straightforward to evaluate the PLQC policy via Monte Carlo simulation. Eventually, it results in a lower bound $V_1^{\widehat{\boldsymbol{\alpha}}}(x_0, f_1)$ on the optimal value function $V_1(x_0, f_1)$.

Remark 2.4.1. *In [62], the authors also consider a time-weighted average price (TWAP) policy where an equal quantity of shares ($a_t = -x_0/T$) is traded at every period, a deterministic (DETER) policy where a deterministic trading strategy is given by solving the deterministic trading problem with all noises ignored, a model predictive control (MPC) policy where the trading strategy is given by solving a sequence of deterministic trading problems without noises, and a linear rebalancing (LRB) policy where the control is given by the best rebalancing policy that is affine in all the available market predictors. Our preliminary numerical tests show that these three policies (TWAP, DETER, MPC) perform significantly worse than the PLQC policy. Although the LRB policy is competitive (often worse) compared with the PLQC policy in terms of performance; it is more sensitive to formulation, more complex in structure, and more difficult to solve. Therefore, we focus on the easy-to-implement PLQC policy.*

2.4.2 Dual Formulation with Regression-based Penalties

Now let us consider the dual formulation of primal problem (2.4.2). Following the derivation in Section 2.2, we know that the dual problem

$$(D):\ V_1^M(x_0, f_1) = \mathbb{E}_1\left[\max_{\boldsymbol{\alpha}\in\mathbb{A}} \sum_{t=1}^{T}\left(x_t^T B f_t - \frac{1}{2}a_t^T \Lambda a_t - \frac{\gamma}{2}x_t^T \Sigma x_t\right) - M(\boldsymbol{\alpha}, \mathbf{z}^{(2)})\right], \quad (2.4.6)$$

where $\mathbf{z}^{(2)} := (z_1^{(2)}, ..., z_T^{(2)})$ and $M(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ is a feasible dual penalty. It follows that $V_1^M(x_0, f_1) \geq V_1(x_0, f_1)$. In particular, when plugging in the optimal dual penalty $M^*(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ the exact value function is recovered, i.e., $V_1^{M^*}(x_0, f_1) = V_1(x_0, f_1)$, where

$$M^*(\boldsymbol{\alpha}, \mathbf{z}^{(2)}) = \sum_{t=1}^{T-1}\Big(V_{t+1}(x_t, f_{t+1}) - \mathbb{E}\left[V_{t+1}(x_t, f_{t+1})|x_{t-1}, a_t, f_t\right]\Big).$$

Since the optimal value functions $\{V_{t+1}(x_t, f_{t+1})\}$ are not available, we have to approximate $M^*(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$. In particular, we will apply the proposed regression approach. Note that the objective function is quadratic, we will use the second-order Taylor series basis (which is equivalent to the second-order L^2 orthonormal basis for this problem) in the algorithm.

Following the derivation in Chapter 2.2.3, let us express the optimal dual penalty w.r.t. the basis $\{d_1 = z, d_2 = z^2 - \mathbb{E}[z^2]\}$, i.e.,

$$M^*(\boldsymbol{\alpha}, \mathbf{z}^{(2)}) \approx \sum_{t=1}^{T-1}\left(\gamma_{t,1}^T(x_{t-1}, a_t, f_t) z_{t+1}^{(2)} + \gamma_{t,2}^T(x_{t-1}, a_t, f_t)\left((z_{t+1}^{(2)})^2 - \mathbb{E}\left[(z_{t+1}^{(2)})^2\right]\right)\right),$$

where $\gamma_{t,r}(x_{t-1}, a_t, f_t)$ are computed via

$$\begin{cases} \gamma_{t,1}(x_{t-1}, a_t, f_t) = Cov(z_{t+1}^{(2)})^{-1}\mathbb{E}_t\left[V_{t+1}(x_t, f_{t+1}) z_{t+1}^{(2)}\right], \\ \gamma_{t,2}(x_{t-1}, a_t, f_t) = Cov((z_{t+1}^{(2)})^2)^{-1}\mathbb{E}_t\left[V_{t+1}(x_t, f_{t+1})\left((z_{t+1}^{(2)})^2 - \mathbb{E}\left[(z_{t+1}^{(2)})^2\right]\right)\right]. \end{cases} \quad (2.4.7)$$

Note that we are slightly abusing the notation $(z_{t+1}^{(2)})^2$ to denote the componentwise square of a random vector. We will use Algorithm 2.2.1 to estimate $\{\gamma_{t,r}(x_{t-1}, a_t, f_t)\}$ in (2.4.7). The key is how to construct proper regressors in the regression. As discussed in the previous section, a good candidate regressor is $\frac{\partial^r}{\partial f^r}\widetilde{V}_{t+1}(x_t, \widehat{f}_{t+1})\cdot(\sigma)^r$, where $\widetilde{V}_{t+1}(\cdot)$ is either the value function of a naive policy or the value function of a

simpler DP such that its derivatives admit closed-form expressions. Here $\widehat{f}_{t+1} = (I - \Phi)f_t$. Recall that the corresponding unconstrained problem is a LQC problem, and its value function $\{J_{t+1}(x_t, f_{t+1})\}$ admit closed-form expressions as in (2.4.3). Therefore, we can use $\frac{\partial^r}{\partial f^r} J_{t+1}(x_t, \widehat{f}_{t+1}) \cdot (\sigma)^r$ as one of regressors for regressing $\gamma_{n,r}(x_{t-1}, a_t, f_t)$. Moreover, $J_{t+1}(x_t, \widehat{f}_{t+1})$ is quadratic in $(x_t, \widehat{f}_{t+1})$, x_t is affine in (x_{t-1}, a_t), and $\widehat{f}_{t+1}$ is linear in f_t. Thus, the first-order and second-order derivatives of $J_{t+1}(x_t, \widehat{f}_{t+1})$ w.r.t. f are both affine in (x_{t-1}, a_t, f_t). It follows that the regressor, and thus the resulted dual penalty is affine in policy and preserves convexity in the dual problem. In particular, our numerical tests show that the regressors

$$\boldsymbol{\phi}_{t,1}(x_{t-1}, a_t, f_t) = \left\{1, \frac{\partial}{\partial f} J_{t+1}(x_t, \widehat{f}_{t+1}) = A^T_{xf,t+1}(x_{t-1} + a_t) + A_{ff,t+1}(I - \Phi)f_t\right\}$$

and

$$\boldsymbol{\phi}_{t,2}(x_{t-1}, a_t, f_t) = \left\{\frac{\partial^2}{\partial f^2} J_{t+1}(x_t, \widehat{f}_{t+1}) = diag(A_{ff,t+1})\right\} \tag{2.4.8}$$

are sufficiently good for regressing $\gamma_{n,1}(x_{t-1}, a_t, f_t)$ and $\gamma_{n,2}(x_{t-1}, a_t, f_t)$, respectively. Here note that the derivatives are componentwise, and $diag(A_{ff,t+1})$ denotes the diagonal vector of matrix $A_{ff,t+1}$. To this end, let us denote the resulted regression-based dual penalty as

$$\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)}) = \sum_{t=1}^{T-1} \left(\widetilde{\gamma}^T_{t,1}(x_{t-1}, a_t, f_t) z^{(2)}_{t+1} + \widetilde{\gamma}^T_{t,2}(x_{t-1}, a_t, f_t) \left((z^{(2)}_{t+1})^2 - \mathbb{E}\left[(z^{(2)}_{t+1})^2\right]\right)\right), \tag{2.4.9}$$

where $\widetilde{\gamma}_{t,r}(x_{t-1}, a_t, f_t)$ is the estimation of $\gamma_{t,r}(x_{t-1}, a_t, f_t)$ after regression. Notice that the regressor $\boldsymbol{\phi}_{t,2}(x_{t-1}, a_t, f_t)$ in (2.4.8) does not depend on (x_{t-1}, a_t, f_t), it follows that $\widetilde{\gamma}^T_{t,2}(x_{t-1}, a_t, f_t)$ does not depend on (x_{t-1}, a_t, f_t). Therefore, the second component of $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$,

$$\sum_{t=1}^{T-1} \widetilde{\gamma}^T_{t,2}(x_{t-1}, a_t, f_t) \left((z^{(2)}_{t+1})^2 - \mathbb{E}\left[(z^{(2)}_{t+1})^2\right]\right)$$

does not involve states and actions. Thus, it does not affect the optimization in the inner optimization problems. In fact, it plays the role of control variate in the sense that it has zero mean and does not affect the expectation of the upper bound

estimator; however, the correlation between this term and the reward function might help reduce its variance. Plugging $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ into the dual problem (2.4.6), we have

$$V_1^{\widetilde{M}^t}(x_0, f_1) = \mathbb{E}_1 \left[\max_{\boldsymbol{\alpha} \in \mathbb{A}} \sum_{t=1}^{T} \left(x_t^T B f_t - \frac{1}{2} a_t^T \Lambda a_t - \frac{\gamma}{2} x_t^T \Sigma x_t \right) - \widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)}) \right].$$

Note that $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ is a feasible dual penalty, and thus $V_1^{\widetilde{M}^t}(x_0, f_1) \geq V_1(x_0, f_1)$. To obtain an unbiased estimator of $V_1^{\widetilde{M}^t}(x_0, f_1)$, we can simulate L i.i.d. sample paths of $\mathbf{z}^{(2)}$, solve the deterministic inner optimization problem corresponding to every sample path, and finally take the sample average of the optimal values. Since the inner optimization problems are deterministic and convex, we use the CVX package (Mosek solver) in Matlab to solve them.

2.4.3 Numerical Results

We will use the model parameters calibrated in [62]. However, the authors only consider the case of trading one stock (Apple, Inc), i.e., $D = 1$. To exhibit the effectiveness of our regression approach for high-dimensional DPs, we "replicate" their model parameters to obtain a high-dimensional dynamic trading problem. The model parameters are summarized as follows.

Time horizon is $T = 12$ or 24. The number of stocks to be liquidated is $D = 1, 5, 10,$ or 25. The initial position of every stock to be liquidated is $[x_0^1, ..., x_0^D]$, a vector in $\mathbb{R}^D$ with $x_0^d = 10000$. There are $K = 2$ market factors to predict the return on each stock, each with a different mean reversion speed. The return of the stocks follows the dynamics

$$r_{t+1}^d = 0.0726 + 0.3375 f_{t,1} - 0.0720 f_{t,2} + z_{t+1,d}^{(1)},\ d = 1, ..., D;\ t = 0, ..., T-1,$$

where $\Sigma = Var(z_{t+1,d}^{(1)}) = 0.048$. Thus, the $D \times K$ matrix

$$B = [0.3375, -0.072; 0.3375, -0.072; ..., 0.3375, -0.072].$$

Similarly, the market factor f follows the dynamics

$$\begin{cases} f_{t+1,1} - f_{t,1} = -0.5 f_{t,1} + z^{(2)}_{t+1,1}, \\ f_{t+1,2} - f_{t,2} = -0.7 f_{t,2} + z^{(2)}_{t+1,2}, \end{cases}$$

where

$$\Psi = Var(z^{(2)}_{t+1}) = \begin{bmatrix} 0.0379 & 0 \\ 0 & 0.0947 \end{bmatrix}.$$

Hence,

$$\Phi = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix}.$$

We also test other choices of Φ in the appendix. Note that we inflate the matrix Φ used in [62] so that the market is more dynamic and the market factors for return prediction depend less on the history. Instead of generating f_0, f_1 randomly, we let $f_0 = [1, 1]'$ and $f_1 = (I - \Phi) f_0$. The transaction cost matrix $\Lambda = \lambda \cdot \widetilde{\Lambda}$, where $\lambda = 2.14 \times 10^{-5}$, and $\widetilde{\Lambda} = \Gamma\Gamma^T$. Here Γ is an upper triangular matrix such that

$$\Gamma = \begin{bmatrix} \frac{1}{\sqrt{D}} & \frac{1}{\sqrt{D}} & \cdots & \frac{1}{\sqrt{D}} \\ 0 & \frac{1}{\sqrt{D-1}} & \cdots & \frac{1}{\sqrt{D-1}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{D \times D}$$

The reason for using such a matrix is to ensure that diagonal elements of $\widetilde{\Lambda}$ are all one, meaning the transaction costs are the same across all the stocks. We also test other choices of λ in the appendix.

With the above model parameters, we solve for the parameters (i.e., A_{xx}, A_{xf}, A_{ff}, and A) of the optimal policy to the corresponding unconstrained problem. Then the PLQC policy is computed via policy projection (2.4.4). To evaluate the PLQC policy and generate a lower bound estimator, we run a simulation of $M = 10^6$ i.i.d. sample paths with the same initial state (x_0, f_1), exercise the PLQC policy along each sample path, and compute the corresponding accumulated reward. The value

function of the PLQC policy, i.e., the lower bound on the optimal value function, is estimated by taking the average of those reward samples.

As for the upper bound estimation, we consider the upper bounds induced by three different feasible dual penalties. In particular, UpperBound 1 is the upper bound generated by the zero penalty (note that zero penalty is a trivial feasible dual penalty). In other words, it is the value function when the investor has a perfect foresight of all the future information without any penalization. UpperBound 2 and UpperBound 3 correspond to the upper bounds generated by the regression-based dual penalties, where UpperBound 2 is induced by using the first-order Taylor series basis, and UpperBound 3 is induced by using the second-order Taylor series basis. That is, UpperBound 2 is induced by the penalty

$$\widetilde{M}_1^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)}) = \sum_{t=1}^{T-1} \widetilde{\gamma}_{t,1}^T(x_{t-1}, a_t, f_t) z_{t+1}^{(2)}, \tag{2.4.10}$$

and UpperBound 3 is induced by the penalty $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ in (2.4.9). To regress $\gamma_{t,r}^T(x_{t-1}, a_t, f_t)$, we implement Algorithm 2.2.1 with the PLQC policy and the $M = 10^6$ sample paths generated in estimating the lower bound. After constructing all three dual penalties, we simulate $L = 400$ sample paths to generate the upper bound estimators. L is much smaller than M because a good approximation of the optimal dual penalty induces an upper bound estimator with a small variance. To illustrate the performance of the PLQC policy, we compute its duality gap, which is the ratio of the difference between the lower bound and the tightest upper bound to the lower bound. If the duality gap is small enough, we could claim the policy is sufficiently good. The detailed results are summarized in Table 2.4.1.

In Table 2.4.1, Column 1 records the number of securities D, including the cases where $D = 1, 5, 10$ or 25. Column 2 records the number of trading horizons T, including the cases $T = 12$ or 24. Column 3 records the lower bounds induced by the PLQC policy. Column 4 records the upper bounds (UpperBound 1) induced by zero

Table 2.4.1: Dual Bounds of PLQC Policy.

D	T	Lower Bound PLQC Policy ($\$k$)	UpperBound 1 Zero Penalty ($\$k$)	UpperBound 2 First-Order ($\$k$)	UpperBound 3 Second-Order ($\$k$)	Duality Gap
1	12	3.652 (0.015)	4.375 (0.225)	3.838 (0.055)	3.854 (0.037)	5.09%
5	12	14.937 (0.074)	18.096 (1.304)	15.365 (0.107)	15.263 (0.055)	2.18%
10	12	24.303 (0.140)	30.240 (2.224)	24.544 (0.152)	24.645 (0.070)	0.99%
25	12	32.090 (0.324)	44.124 (5.237)	32.657 (0.211)	32.529 (0.109)	1.37%
1	24	3.901 (0.019)	6.317 (0.397)	5.206 (0.174)	5.224 (0.171)	33.91%
5	24	18.635 (0.103)	28.060 (2.094)	20.012 (0.250)	20.103 (0.152)	7.39%
10	24	33.309 (0.212)	47.738 (3.816)	35.099 (0.306)	35.285 (0.193)	5.37%
25	24	62.971 (0.528)	86.537 (9.673)	64.401 (0.370)	65.224 (0.201)	2.27%

penalty. Column 5 records the upper bounds (UpperBound 2) induced by the dual penalty $\widetilde{M}_1^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ in (2.4.10). Column 6 records the upper bounds (UpperBound 3) induced by the dual penalty $\widetilde{M}^t(\boldsymbol{\alpha}, \mathbf{z}^{(2)})$ in (2.4.9). The half confidence interval widths of the lower bound and each upper bound are presented in the parentheses. The last column records the best duality gaps achieved by comparing the lower bounds and the tightest upper bounds. We have the following observations:

- In general, in view of the small duality gaps in most cases, we could claim that the simple PLQC policy is a sufficiently good policy.

- Comparing UpperBound 1, UpperBound 2 and UpperBound 3 induced by the zero penalty, the first-order regression-based penalty, and the second-order regression-based penalty, respectively, we notice: 1) UpperBound 1 have large duality gaps, meaning zero penalty performs poorly as one would expect. 2) UpperBound 2 and UpperBound 3 have very small duality gaps, meaning that the policy used in the regression (here is the PLQC policy) is near optimal and the regression-based dual penalties are accurate approximations of the optimal dual

penalty. 3) It is hard to distinguish between UpperBound 2 and UpperBound 3 in terms of tightness. The reason is that the inner optimization problems are the same, except for a control variate term in the second-order regression-based penalty. This term helps reduce the variance of the upper bound estimator, which is verified by the fact that the half confidence interval widths of UpperBound 3 are significantly narrower than the ones of UpperBound 2.

- We also observe that, as the number of trading horizons T increases, the duality gaps increase. It might imply that the PLQC policy becomes less optimal for larger T. Another possibility is that the approximation of the optimal dual penalty is less accurate for larger T.

- Lastly, as the number of securities in position increases, the duality gaps decrease instead of increasing as one would expect. One possible explanation is that the PLQC policy is closer to optimal for higher-dimensional problems due to the strong correlations across the assets in the transaction cost matrix Λ. Therefore, the optimal trading strategy of the unconstrained problem is more conservative, and thus less optimality is lost in the policy projection.

Overall, we can see that the PLQC policy is near optimal for the model parameters tested. More importantly, the regression-based dual penalties perform well. It implies the effectiveness and efficiency of the proposed framework of regression approach in solving duals of high-dimensional DPs.

2.5 Conclusion

In this chapter, we develop a framework of regression approach to approximating the optimal dual penalty in general DPs, by studying the structure of the dual penalty space. The proposed framework circumvents the issue of nested simulation suffered

by some of the existing approaches, and thus improves computational efficiency. Furthermore, it requires minimal extra simulation and computational costs by reusing the samples generated during the estimation of lower bound. Lastly, it is very robust. Dual penalties with desired structural properties could be generated by constructing proper regressors in the regression. The application to a high-dimensional dynamic trading problem demonstrates its effectiveness in generating good feasible dual penalties and tight upper bounds on the optimal value function.

CHAPTER III

RISK ASSESSMENT IN STOCHASTIC SIMULATION UNDER INPUT UNCERTAINTY

In this chapter, we will present the framework of risk assessment of stochastic simulation under input uncertainty, and we will study the research goals proposed in 1.2.1. This work is also summarized in our paper [91].

The rest of the chapter is organized as follows. In Chapter 3.1, we introduce risk measures such as VaR and CVaR, and the corresponding nested risk estimators for risk assessment of stochastic simulation under input uncertainty. In Chapter 3.2, we present the asymptotical properties of nested risk estimators, and describe the construction of (asymptotically valid) confidence intervals. We study the associated budget allocation problem in Chapter 3.3. In Chapter 3.4, we conduct numerical experiments to demonstrate the theoretical results in previous sections. Conclusions and promising directions of future research are provided in Chapter 3.5.

3.1 Risk Measures of Mean Response

3.1.1 A Risk Formulation

For a stochastic system, loosely speaking, risk assessment of the system under input uncertainty characterizes the extreme behavior of mean system response under all possible input model scenarios. Therefore, it is of great importance in real-world system design/control, especially when extreme mean response might lead to disastrous outcomes. Risk measures of mean response w.r.t. input uncertainty provide rigorous quantifications of such extreme behavior, and thus are excellent candidates for risk assessment under input uncertainty. In this section, we will introduce general risk

measures under input uncertainty and their estimation. In particular, we will focus on analyzing nested Monte Carlo estimators of two widely used risk measures—VaR and CVaR of mean response w.r.t input uncertainty.

Let us first rigorously define the risk measures VaR and CVaR of mean response w.r.t. input uncertainty. In a stochastic simulation experiment, consider an output response function in the form of $h(\theta;\xi)$, where θ represents the input parameter(s) and ξ represents the noise (stochastic uncertainty) in the response. Furthermore, suppose there is a probability distribution (so-called "belief distribution") on θ that reflects our belief on input uncertainty, since θ needs to be inferred from finite historical data. For example, if one takes a Bayesian approach, the belief distribution on input parameter will be updated with new observations. Specifically, suppose there exists a prior distribution $p(\theta)$ on θ, it could be either non-informative or informative depending on prior knowledge, and the hyper-parameters of the prior could be estimated from historical data. With new observations $\mathbf{x}$, a posterior distribution $p(\cdot|\mathbf{x})$ on θ is obtained via standard Bayesian updating. Of course, there are other approaches such as Bootstrapping to construct the belief distribution on θ. We further denote

$$h(\theta;\xi) = H(\theta) + \mathcal{E}(\theta;\xi),$$

where $H(\theta) = \mathbb{E}_\xi[h(\theta;\xi)]$ is the mean response, and $\mathcal{E}(\theta;\xi)$ is the stochastic noise that satisfies $\mathbb{E}[\mathcal{E}(\theta;\xi)|\theta] = 0$ and $Var[\mathcal{E}(\theta;\xi)|\theta] = \tau_\theta^2$, where τ_θ^2 is a finite deterministic function of θ and $\tau^2 \stackrel{\triangle}{=} \int \tau_\theta^2 p(\theta|\mathbf{x})d\theta$ is finite. Furthermore, to account for input uncertainty, let $H(\theta) = H^{\mathbf{x}} + \mathcal{Z}(\theta)$, where $\mathcal{Z}(\theta)$ represents the error of mean response due to input uncertainty that satisfies $\mathbb{E}_{p(\cdot|\mathbf{x})}[\mathcal{Z}(\theta)] = 0$ and $Var_{p(\cdot|\mathbf{x})}[\mathcal{Z}(\theta)] = \sigma^2$, where σ^2 depends on historical data $\mathbf{x}$.

Assume $0 < \alpha < 1$ is a certain large probability level of interest (usually, $\alpha = 0.95$ or $\alpha = 0.99$). The Value-at-Risk $VaR_\alpha\left(\mathbb{E}_\xi[h(\theta;\xi)]\right)$ (or interchangeably $VaR_\alpha\left(H(\theta)\right)$)

is defined by the α-quantile of mean response $H(\theta)$, i.e.,

$$VaR_\alpha\left(H(\theta)\right) := \inf\{t : F(t) \geq \alpha\}, \tag{3.1.1}$$

where $F(\cdot)$ is the cumulative distribution function (c.d.f.) of $H(\theta)$. When $H(\theta)$ admits a continuous and positive probability density function (p.d.f.) $f(\cdot)$ around $VaR_\alpha\left(H(\theta)\right)$, definition (3.1.1) can be simplified as $VaR_\alpha\left(H(\theta)\right) = F^{-1}(\alpha)$. Intuitively, $VaR_\alpha\left(H(\theta)\right)$ represents the cut-off value for the α-tail of mean response. CVaR, on the other hand, is defined by the conditional expectation of the α-tail of the mean response distribution. Under previous assumption, $CVaR_\alpha\left(\mathbb{E}_\xi[h(\theta;\xi)]\right)$ (or interchangeably $CVaR_\alpha\left(H(\theta)\right)$) is defined by

$$CVaR_\alpha\left(\mathbb{E}_\xi[h(\theta;\xi)]\right) := \frac{1}{1-\alpha}\mathbb{E}_{p(\cdot|\mathbf{x})}\Big[\mathbb{E}_\xi[h(\theta;\xi)]\mathbb{1}\{\mathbb{E}_\xi[h(\theta;\xi)] \geq VaR_\alpha\}\Big], \tag{3.1.2}$$

where the indicator function $\mathbb{1}\{A\}$ equals 1 when statement A is true and 0 otherwise. With slight abuse of notations, we use VaR_α as an abbreviation for $VaR_\alpha\left(H(\theta)\right)$ and $CVaR_\alpha$ as an abbreviation for $CVaR_\alpha\left(H(\theta)\right)$, respectively.

Computing risk measures such as VaR and CVaR w.r.t. input uncertainty could be straightforward when the system is simple. For example, when there is no stochastic uncertainty and the response (as a random variable induced by input uncertainty) admits an explicit p.d.f., VaR/CVaR of the response could be easily computed via numerical integration.

However, for complex real-world stochastic systems, computing risk measures directly is generally not applicable. In this case, Monte Carlo simulation is a powerful alternative approach to obtain good estimate of VaR/CVaR. Next, we will describe nested Monte Carlo estimators for VaR and CVaR of mean response.

3.1.2 Nested Monte Carlo Estimators

VaR and CVaR, as two of the mostly used risk measures in financial applications, have been extensively studied. Estimation of VaR and CVaR are of great interest,

and they could be quite challenging problems in the area of rare-event simulation. To gain more intuition, let us first consider the estimation of VaR_α and $CVaR_\alpha$ without the presence of stochastic uncertainty. That is, $H(\theta)$ can be evaluated exactly given any θ. Now that the probabilistic model is one layer, a natural approach to estimating VaR_α and $CVaR_\alpha$ is by naive Monte Carlo sampling described as follows. First, draw N i.i.d. scenarios $\theta_1, ..., \theta_N$ from the belief distribution $p(\theta|\mathbf{x})$; second, evaluate the response $H(\theta_i)$ for $i = 1, ..., N$ and sort the resulting response scenarios $H(\theta_1), ..., H(\theta_N)$ in ascending order, denoted by $H(\theta_{(1)}) \leq H(\theta_{(2)}) \leq \cdots \leq H(\theta_{(N)})$; finally, the Monte Carlo estimators of VaR_α and $CVaR_\alpha$ are given respectively by

$$\begin{cases} \widehat{VaR}_\alpha = H(\theta_{(\alpha N)}), \\ \widehat{CVaR}_\alpha = \frac{1}{(1-\alpha)N} \sum_{i=1}^{N} H(\theta_i) \mathbb{1}\{H(\theta_i) \geq H(\theta_{(\alpha N)})\} = \frac{1}{(1-\alpha)N} \sum_{i=\alpha N}^{N} H(\theta_{(i)}), \end{cases} \tag{3.1.3}$$

where we assume αN is an integer for convenience. Intuitively, $\widehat{VaR}_\alpha$ is the α-quantile of the empirical distribution of response scenarios, i.e., the α-level VaR of the empirical distribution. Similarly, $\widehat{CVaR}_\alpha$ is the average of response scenarios that are greater than or equal to $\widehat{VaR}_\alpha$, where $\widehat{VaR}_\alpha$ functions as an approximation of VaR_α. The properties of these two estimators have been well-studied. For example, although $\widehat{VaR}_\alpha$ and $\widehat{CVaR}_\alpha$ are biased, they have been proven to be strongly consistent and asymptotically normally distributed by [80] under appropriate regularity conditions.

Now back to the original VaR/CVaR estimation problem that we are interested in, where stochastic uncertainty needs to be taken into account as well. Because the mean response $H(\theta)$ is not directly assessable, it is estimated from samples. To obtain estimators of VaR_α and $CVaR_\alpha$, we can extend the sampling procedure described previously by replacing $\{H(\theta_i)\}$ with the corresponding sample average estimates $\{\widehat{H}(\theta_i)\}$. Specifically, for $i = 1, ..., N$, draw M i.i.d. samples $\xi_{i1}, ..., \xi_{iM}$ from the distribution of ξ and evaluate the responses $h(\theta_i; \xi_{ij}), j = 1, ..., M$; approximate the mean response $H(\theta_i)$ by $\widehat{H}_M(\theta_i) = \frac{1}{M} \sum_{j=1}^{M} h(\theta_i; \xi_{ij})$ and sort them in ascending order, denoted by $\widehat{H}_M(\theta^{(1)}) \leq \widehat{H}_M(\theta^{(2)}) \leq \cdots \leq \widehat{H}_M(\theta^{(N)})$. Here note that $(\theta^{(1)}, ..., \theta^{(N)})$

and $(\theta_{(1)}, ..., \theta_{(N)})$ are different order statistics due to inner sampling. In fact, for fixed scenarios $\theta_1, ..., \theta_N$, $(\theta_{(1)}, ..., \theta_{(N)})$ is a constant vector, while $(\theta^{(1)}, ..., \theta^{(N)})$ is its random permutation that depends on realizations of $h(\theta_i; \xi_{ij})$'s. Finally, the nested estimators of VaR_α and $CVaR_\alpha$ are given respectively by

$$\begin{cases} \widetilde{VaR}_\alpha = \widehat{H}_M(\theta^{(\alpha N)}), \\ \widetilde{CVaR}_\alpha = \frac{1}{(1-\alpha)N} \sum_{i=1}^{N} \widehat{H}_M(\theta_i) \mathbb{1}\{\widehat{H}_M(\theta_i) \geq \widehat{H}_M(\theta^{(\alpha N)})\} = \frac{1}{(1-\alpha)N} \sum_{i=\alpha N}^{N} \widehat{H}_M(\theta^{(i)}). \end{cases} \tag{3.1.4}$$

With the complication of stochastic uncertainty, the asymptotical properties of $\widetilde{VaR}_\alpha$ and $\widetilde{CVaR}_\alpha$ become more difficult to analyze. In next section, we will show that the estimators maintain to be strongly consistent and asymptotically normally distributed in different limiting senses under different sets of regularity conditions. Hence, using them as inferences for VaR_α and $CVaR_\alpha$ respectively is still reasonable.

Remark 3.1.1. *[7] and [82] essentially use the nested VaR estimator in* (3.1.4) *(at specified* α*) as the boundary of their confidence interval (CI) or credible interval (CrI) for* $H(\theta_c)$*—the mean response at the true but unknown input parameter* θ_c*, so that the structural bias of using* $\frac{1}{N}\sum_{i=1}^{N} \widehat{H}_M(\theta_i)$ *as the estimator of* $H(\theta_c)$ *(because* $\mathbb{E}_\theta[H(\theta)] \neq H(\theta_c)$*) is covered by the CI or CrI. In this chapter, we aim to quantify the extreme mean response w.r.t. the belief distribution of input parameter. We are not aiming to provide inferences on the risk of response at the true but unknown input parameter* θ_c*, such as* $VaR_\alpha(h(\theta^c, \xi))$*, in the presence of input uncertainty. Therefore, there is no structural bias between the proposed nested risk estimators and the quantities of interest.*

3.2 Asymptotical Analysis

3.2.1 Consistency

In this subsection, we analyze the asymptotical properties of $\widetilde{VaR}_\alpha$ and $\widetilde{CVaR}_\alpha$ in (3.1.4) as both the outer and the inner sample sizes go to infinity. In particular, we will prove the strong consistency of nested risk estimators in different limiting

senses under different sets of regularity assumptions. We refer to them as "*Weak Assumption*" and "*Strong Assumption*", respectively.

Assumption 3.2.1. ***(Weak)*** *(i). The response $h(\theta;\xi)$ has finite conditional second moment, i.e., $\tau_\theta^2 := \mathbb{E}[h^2(\theta;\xi)|\theta] < \infty$ w.p.1, and $\tau^2 := \int \mathbb{E}[h^2(\theta;\xi)|\theta] f(\theta) d\theta < \infty$.*
(ii). The mean response $H(\theta)$ is a continuous random variable. Furthermore, the p.d.f. of $H(\theta)$, $f(\cdot)$ is positive and continuously differentiable around VaR_α.

Assumption 3.2.1 is referred to as "Weak Assumption", because it imposes separate assumptions on the distribution of mean response $H(\theta)$ due to input uncertainty and on the distribution of $h(\theta;\xi)$ due to stochastic uncertainty. Unlike the following "Strong Assumption", it does not impose assumptions on the joint distribution of input uncertainty and stochastic uncertainty.

Notice that

$$\widehat{H}_M(\theta) \stackrel{\triangle}{=} \frac{1}{M}\sum_{j=1}^{M} h(\theta;\xi_j) = \frac{1}{M}\sum_{j=1}^{M}(H(\theta)+\mathcal{E}(\theta;\xi_j)) = H(\theta) + \frac{1}{M}\sum_{j=1}^{M}\mathcal{E}(\theta;\xi_j).$$

Let us denote the "normalized" noise $\bar{\mathcal{E}}_M \stackrel{\triangle}{=} \sqrt{M}\cdot\frac{1}{M}\sum_{j=1}^{M}\mathcal{E}(\theta;\xi_j)$ (so that $\bar{\mathcal{E}}_M$ has a limiting distribution as the inner sample size $M\to\infty$ under appropriate assumptions), and thus $\widehat{H}_M(\theta) = H(\theta) + \bar{\mathcal{E}}_M/\sqrt{M}$. We further denote by $\widetilde{f}_M(\cdot)$ and $\widetilde{F}_M(\cdot)$ the p.d.f. and c.d.f. of $\widehat{H}_M(\theta)$, respectively. When $\bar{\mathcal{E}}_M$ has a limiting distribution, the "distance" between the distribution of $\widehat{H}_M(\theta)$ and the distribution of $H(\theta)$ vanishes as $M\to\infty$, i.e., $\widetilde{f}_M \to f$ and $\widetilde{F}_M \to F$ as $M\to\infty$. Intuitively, the following Strong Assumption guarantees that $\widetilde{f}_M$ and $\widetilde{F}_M$ converge sufficiently fast.

Assumption 3.2.2. ***(Strong)*** *(i) The response $h(\theta;\xi)$* has finite conditional second moment, i.e., $\tau_\theta^2 := \mathbb{E}[h^2(\theta;\xi)|\theta] < \infty$ w.p.1 and $\tau^2 := \int \mathbb{E}[h^2(\theta;\xi)|\theta] f(\theta) d\theta < \infty$.
(ii) The joint density $p_M(h,e)$ of $H(\theta)$ and $\bar{\mathcal{E}}_M$, and its partial derivatives $\frac{\partial}{\partial h}p_M(h,e)$ and $\frac{\partial^2}{\partial h^2}p_M(h,e)$ exist for each M and for all pairs of (h,e).
(iii) There exist nonnegative functions $g_{0,M}(\cdot)$, $g_{1,M}(\cdot)$ and $g_{2,M}(\cdot)$ such that $p_M(h,e) \le$

$g_{0,M}(e)$, $\left|\frac{\partial}{\partial h}p_M(h,e)\right| \leq g_{1,M}(e)$, $\left|\frac{\partial^2}{\partial h^2}g_M(h,e)\right| \leq g_{2,M}(e)$ for all h, e. Furthermore, $\sup_M \int |e|^r g_{i,M}(e)de < \infty$ for $i = 0, 1, 2$, and $0 \leq r \leq 4$.

Strong Assumption is stronger than Weak Assumption. In Strong Assumption, (i) ensures that $\bar{\mathcal{E}}_M$ has a limiting distribution as $M \to \infty$, and (ii) and (iii) (which are the assumptions in Assumption 1 of [40]) on the joint distribution of mean response $H(\theta)$ and the "normalized" noise $\bar{\mathcal{E}}_M$ so that the distance between $\widetilde{f}_M(\cdot)$ and $f(\cdot)$ is of $O(\frac{1}{M})$. Strong Assumption is expected to hold when the response function $h(\cdot,\cdot)$ is sufficiently smooth and the distributions of θ and ξ have good structural properties (e.g., finite moments up to some order).

Theorem 3.2.1. *(**Weak Result on Consistency**) Under Assumption 3.2.1, we have*

$$\lim_{N\to\infty}\lim_{M\to\infty}\widetilde{VaR}_\alpha = VaR_\alpha, \quad w.p.1, \tag{3.2.1}$$

$$\lim_{N\to\infty}\lim_{M\to\infty}\widetilde{CVaR}_\alpha = CVaR_\alpha, \quad w.p.1. \tag{3.2.2}$$

The proof is presented in Appendix B.1. The results in Theorem 3.2.1 are "weak" because the limits on the outer sample size N and the inner sample size M in (3.2.1) and (3.2.2) are iterated and non-interchangeable. Intuitively, the inner sample size M needs to go to infinity first to ensure that, for any fixed θ, $\widehat{H}_M(\theta) \to H(\theta)$ w.p.1 (by Strong Law of Large Numbers (SLLN) under Assumption 3.2.1.(i)). It follows that for fixed $\theta_1, ..., \theta_N$, the random order statistics $(\theta^{(1)}, ..., \theta^{(N)}) \to (\theta_{(1)}, ..., \theta_{(N)})$ w.p.1. as $M \to \infty$. Thus, for fixed $\theta_1, ..., \theta_N$, $(\widehat{H}_M(\theta^{(1)}), ..., \widehat{H}_M(\theta^{(N)})) \to (H(\theta_{(1)}), ..., H(\theta_{(N)}))$ w.p.1. as $M \to \infty$, which immediately results in that for fixed $\theta_1, ..., \theta_N$, $\widetilde{VaR}_\alpha \to \widehat{VaR}_\alpha$ and $\widetilde{CVaR}_\alpha \to \widehat{CVaR}_\alpha$ w.p.1 as $M \to \infty$. Next, as the outer sample size $N \to \infty$, one could show that $\widehat{VaR}_\alpha \to VaR_\alpha$ and $\widehat{CVaR}_\alpha \to CVaR_\alpha$ w.p.1.

When Strong Assumption is imposed, one would naturally expect the consistency results in Theorem 3.2.1 to be strengthened. In particular, the following theorem shows that the iterated limits on N and M in Theorem 3.2.1 can be relaxed such that

nested risk estimators $\widetilde{VaR}$ and $\widetilde{CVaR}$ maintain to be strongly consistent when N and M go to infinity simultaneously.

Theorem 3.2.2. *(**Strong Result on Consistency**) Under Assumption 3.2.2, we have*

$$\lim_{N,M\to\infty} \widetilde{VaR}_\alpha = VaR_\alpha, \quad w.p.1, \tag{3.2.3}$$

$$\lim_{N,M\to\infty} \widetilde{CVaR}_\alpha = CVaR_\alpha, \quad w.p.1. \tag{3.2.4}$$

The proof is presented in Appendix B.2. The results in Theorem 3.2.2 are strong in the sense that the limits on N and M are simultaneous limits instead of iterated limits. The intuition is as follows. In showing the weak results in Theorem 3.2.1, we first fix the outer sample size N and study the asymptotical properties of risk estimators as the inner sample size M goes to infinity. To show the strong results in Theorem 3.2.2, we take a reverse approach, meaning that first fixing the inner sample size M and studying the distance between $VaR_\alpha(\widehat{H}_M(\theta))/CVaR_\alpha(\widehat{H}_M(\theta))$ and $VaR_\alpha(H(\theta))/CVaR_\alpha(H(\theta))$, where note that $VaR_\alpha(\widehat{H}_M(\theta))/CVaR_\alpha(\widehat{H}_M(\theta))$ is the exact VaR/CVaR of the "noised" mean response $\widehat{H}_M(\theta)$. As mentioned previously, Assumption 3.2.2 ensures that the distance between $\widetilde{f}_M(\cdot)$ and $f(\cdot)$ is of $O(\frac{1}{M})$. It follows that the difference between $VaR_\alpha(\widehat{H}_M(\theta))/CVaR_\alpha(\widehat{H}_M(\theta))$ and $VaR_\alpha(H(\theta))/CVaR_\alpha(H(\theta))$ is of $O(\frac{1}{M})$ as well. Second, note that $\widetilde{VaR}_\alpha/\widetilde{CVaR}_\alpha$ could be regarded as an one layer VaR/CVaR estimator of $\widehat{H}_M(\theta)$, i.e.,

$$\widetilde{VaR}_\alpha(H(\theta)) = \widehat{VaR}_\alpha(\widehat{H}_M(\theta)) \quad \text{and} \quad \widetilde{CVaR}_\alpha(H(\theta)) = \widehat{CVaR}_\alpha(\widehat{H}_M(\theta)). \tag{3.2.5}$$

Under Assumption 3.2.2, we could show that $\widehat{VaR}_\alpha(\widehat{H}_M(\theta))/\widehat{CVaR}_\alpha(\widehat{H}_M(\theta))$ converges to $VaR_\alpha(\widehat{H}_M(\theta))/CVaR_\alpha(\widehat{H}_M(\theta))$ uniformly for all M as $N \to \infty$, which implies that N and M can go to infinity simultaneously in Theorem 3.2.2.

3.2.2 Asymptotic Normality

After showing the strong consistency of nested risk estimators, it is natural to consider their asymptotic normality properties and construct CIs for the estimators. Corresponding to the weak/strong results on consistency, the weak/strong normality results for the estimators could be established under Weak Assumption and Strong Assumption, respectively.

First, let us investigate the estimators' weak normality under Weak Assumption. Following the logics in Theorem 3.2.1's proof (see Appendix B.1), the total error of nested risk estimators is decomposed into two components that are caused by input uncertainty and stochastic uncertainty, respectively. Specifically,

$$\widetilde{VaR}_\alpha - VaR_\alpha = \left(\widetilde{VaR}_\alpha - \widehat{VaR}_\alpha\right) + \left(\widehat{VaR}_\alpha - VaR_\alpha\right) \stackrel{\triangle}{=} Err_1 + Err_2, \quad (3.2.6)$$

and

$$\widetilde{CVaR}_\alpha - CVaR_\alpha = \left(\widetilde{CVaR}_\alpha - \widehat{CVaR}_\alpha\right) + \left(\widehat{CVaR}_\alpha - CVaR_\alpha\right) \stackrel{\triangle}{=} Err_3 + Err_4, \quad (3.2.7)$$

where Err_1/Err_3 is caused by stochastic uncertainty and Err_2/Err_4 is caused by input uncertainty. Furthermore, Err_1 and Err_2 are correlated, and so are Err_3 and Err_4. Therefore, it is natural to establish asymptotic normality results and construct CIs for Err_1/Err_3 and Err_2/Err_4, respectively; and then integrate the two resulting CIs to obtain a final CI for nested risk estimators. Of course, as a two-level CI procedure, the relaxation from applying Boole's Inequality results in a wider CI, as general two-level CI procedures do. Therefore, in practice the final CI usually achieves a coverage probability that is much greater than the target confidence level (as shown in the first numerical experiment in Chapter 3.4). To this end, the following theorem establishes the weak asymptotic normality for nested risk estimators and provides an explicit characterization of (asymptotic) variances.

Theorem 3.2.3. *(**Weak Normality**) Under Assumption 3.2.1, we have*

$$\lim_{N\to\infty} \sqrt{N}\left(\widehat{VaR}_\alpha - VaR_\alpha\right) \stackrel{\mathcal{D}}{\Rightarrow} \sigma_{var}\mathcal{N}(0,1), \quad (3.2.8)$$

$$\lim_{N\to\infty} \sqrt{N}\left(\widehat{CVaR}_\alpha - CVaR_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{cvar}\mathcal{N}(0,1), \tag{3.2.9}$$

where

$$\sigma_{var} \overset{\triangle}{=} \frac{\sqrt{\alpha(1-\alpha)}}{f(VaR_\alpha)} \quad \textit{and} \quad \sigma_{cvar} \overset{\triangle}{=} \frac{\sqrt{Var\left[(H(\theta)-VaR_\alpha)^+\right]}}{(1-\alpha)}.$$

Furthermore,

$$\lim_{N\to\infty}\lim_{M\to\infty} \sqrt{M}\left(\widetilde{VaR}_\alpha - \widehat{VaR}_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \tau_{var}\mathcal{N}(0,1), \tag{3.2.10}$$

$$\lim_{N\to\infty}\lim_{M\to\infty} \sqrt{(1-\alpha)NM}\left(\widetilde{CVaR}_\alpha - \widehat{CVaR}_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \tau_{cvar}\mathcal{N}(0,1), \tag{3.2.11}$$

where

$$\tau_{var} \overset{\triangle}{=} \sqrt{\mathbb{E}[\tau_\theta^2 | H(\theta) = VaR_\alpha]} \quad \textit{and} \quad \tau_{cvar} \overset{\triangle}{=} \sqrt{\mathbb{E}[\tau_\theta^2 | H(\theta) \geq VaR_\alpha]}.$$

The proof is presented in Appendix B.3. Assume β is the target error level (hence $1-\beta$ is the target confidence level), let us construct two-sided CIs for the estimators with confidence level at least $1-\beta$. Following the error decomposition, the error level β is decomposed into β_O and β_I (hence $\beta = \beta_O + \beta_I$) as well, representing the error levels for input uncertainty (outer-layer simulation) and stochastic uncertainty (inner-layer simulation), respectively.

Specifically, by (3.2.8) and (3.2.9), the two-sided (unknown variance) CIs for $\widehat{VaR}_\alpha - VaR_\alpha$ and $\widehat{CVaR}_\alpha - CVaR_\alpha$ with confidence level $1-\beta_O$ are

$$\widehat{VaR}_\alpha - VaR_\alpha \in \left[\frac{t_{\beta_O/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}}, \frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}}\right], \tag{3.2.12}$$

$$\widehat{CVaR}_\alpha - CVaR_\alpha \in \left[\frac{t_{\beta_O/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}}, \frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}}\right], \tag{3.2.13}$$

where $\widehat{\sigma}_{var}$ is the sample estimate of $\sigma_{var} \overset{\triangle}{=} \sqrt{\alpha(1-\alpha)}/f(VaR_\alpha)$, in which $f(VaR_\alpha)$ can be estimated using Gaussian kernel density estimation (see, e.g., [79]); $\widehat{\sigma}_{cvar}$ is the sample estimate of $\sigma_{cvar} \overset{\triangle}{=} \sqrt{Var\left[(H(\theta)-VaR_\alpha)^+\right]}/(1-\alpha)$, and $t_{\gamma,L}$ represents the γ-level quantile of a t-distribution with degree of freedom L.

Similarly, by (3.2.10) and (3.2.11), the two-sided (unknown variance) CI for $\widetilde{VaR}_\alpha - \widehat{VaR}_\alpha$ and $\widetilde{CVaR}_\alpha - \widehat{CVaR}_\alpha$ with confidence level $1-\beta_I$ are

$$\widetilde{VaR}_\alpha - \widehat{VaR}_\alpha \in \left[\frac{t_{\beta_I/2,M-1}\widehat{\tau}_{var}}{\sqrt{M}}, \frac{t_{1-\beta_I/2,M-1}\widehat{\tau}_{var}}{\sqrt{M}}\right], \tag{3.2.14}$$

and

$$\widetilde{CVaR}_\alpha - \widehat{CVaR}_\alpha \in \left[\frac{t_{\beta_I/2,(1-\alpha)NM-1}\widehat{\tau}_{cvar}}{\sqrt{(1-\alpha)NM}}, \frac{t_{1-\beta_I/2,(1-\alpha)NM-1}\widehat{\tau}_{cvar}}{\sqrt{(1-\alpha)NM}}\right], \tag{3.2.15}$$

where $\widehat{\tau}_{var}$ and $\widehat{\tau}_{cvar}$ are the sample estimates of τ_{var} and τ_{cvar}, respectively.

Finally, by integrating the CIs in (3.2.12) and (3.2.14), the two-sided (unknown variance) CI for VaR_α with confidence level at least $1-\beta$ is

$$\begin{aligned} VaR_\alpha \in \Bigg[&\widehat{H}_M(\theta^{(\alpha N)}) + \frac{t_{\beta_O/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}} + \frac{t_{\beta_I/2,M-1}\widehat{\tau}_{var}}{\sqrt{M}}, \\ &\widehat{H}_M(\theta^{(\alpha N)}) + \frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}} + \frac{t_{1-\beta_I/2,M-1}\widehat{\tau}_{var}}{\sqrt{M}}\Bigg]. \end{aligned} \tag{3.2.16}$$

Similarly, by integrating the CIs in (3.2.13) and (3.2.15), the two-sided (unknown variance) CI for $CVaR_\alpha$ with confidence level at least $1-\beta$ is

$$\begin{aligned} CVaR_\alpha \in \Bigg[&\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}\widehat{H}_M\theta^{(i)}) + \frac{t_{\beta_O/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}} + \frac{t_{\beta_I/2,(1-\alpha)NM-1}\widehat{\tau}_{cvar}}{\sqrt{(1-\alpha)NM}}, \\ &\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}\widehat{H}_M(\theta^{(i)}) + \frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}} + \frac{t_{1-\beta_I/2,(1-\alpha)NM-1}\widehat{\tau}_{cvar}}{\sqrt{(1-\alpha)NM}}\Bigg] \end{aligned} \tag{3.2.17}$$

We refer to them as "weak CIs". Their structure indicates that we can appropriately choose the outer error level β_O and the inner error level β_I, as well as the outer sample size N and the inner sample size M, to separately assess/control the simulation errors due to input uncertainty and stochastic uncertainty, respectively.

Under Strong Assumption, the asymptotic normality results of nested risk estimators are simpler in form. Following the error decomposition in Appendix B.2, the error of nested risk estimators is decomposed into two components that account for the one layer simulation error due to input uncertainty and the simulation bias due to stochastic uncertainty (see (B.2.2) in Appendix B.2), respectively. By properly

choosing the outer sample size N and the inner sample size M, we can force the bias component to be asymptotically insignificant compared with the one layer simulation error. Specifically, we have the following theorem on the asymptotic normality of nested risk estimators and characterization of (asymptotic) variances.

Theorem 3.2.4. *(**Strong Normality**) Under Assumption 3.2.2, $N = o_M(M^2)$ is the sufficient and necessary condition for*

$$\lim_{N,M\to\infty} \sqrt{N}\left(\widetilde{VaR}_\alpha - VaR_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{var}\mathcal{N}(0,1), \tag{3.2.18}$$

$$\lim_{N,M\to\infty} \sqrt{N}\left(\widetilde{CVaR}_\alpha - CVaR_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{cvar}\mathcal{N}(0,1), \tag{3.2.19}$$

where $N = o_M(M^2)$ means $\lim_{M\to\infty} N/M^2 = 0$.

The proof is presented in Appendix B.4. The results in Theorem 3.2.4 are consistent with the results in [40] on the characterization of asymptotic variances of nested risk estimators. Nevertheless, Theorem 3.2.4 is stronger in the sense that it directly leads to the results in [40]. Moreover, [40] shows that, when minimizing MSE of nested risk estimators, the variance and bias are well-balanced when the sample size pair (N, M) lives in the region of $N = O_M(M^2)$. In contrast, Theorem 3.2.4 shows that nested risk estimators are asymptotically normally distributed when the sample size pair (N, M) lives in the region of $N = o_M(M^2)$.

Following Theorem 3.2.4, we can construct two-sided (unknown variance) CIs for VaR_α and $CVaR_\alpha$ with confidence level at least $1-\beta$, respectively. Specifically,

$$VaR_\alpha \in \left[\widehat{H}_M(\theta^{(\alpha N)}) + \frac{t_{\beta/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}},\ \widehat{H}_M(\theta^{(\alpha N)}) + \frac{t_{1-\beta/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}}\right] \tag{3.2.20}$$

and

$$\begin{aligned} CVaR_\alpha \in \Bigg[&\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}\widehat{H}_M(\theta^{(i)}) + \frac{t_{\beta/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}}, \\ &\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}\widehat{H}_M(\theta^{(i)}) + \frac{t_{1-\beta/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}}\Bigg]. \end{aligned} \tag{3.2.21}$$

Note that the CIs in (3.2.20) and (3.2.21) only depend on the outer sample size N. The reason is that, the $O(\frac{1}{M})$ bias term (see Lemma B.2.2 and B.2.3 in Appendix B.2 for explicit formulas) due to stochastic uncertainty is asymptotically dominated by the $O(\frac{1}{\sqrt{N}})$ deviation terms in (3.2.20) and (3.2.21) when $N = o_M(M^2)$. Nevertheless, in practice, it is better to incorporate the $O(\frac{1}{M})$ bias terms in the CIs so that the one layer simulation error due to input uncertainty and the simulation bias due to stochastic uncertainty could be separately assessed/controlled. We refer to the CIs in (3.2.20) and (3.2.21) with biases taken into account as "strong CIs".

The similarity between weak and strong CIs is that they both consist of two components caused by input uncertainty and stochastic uncertainty respectively, and could be assessed/controlled separately. The difference between weak and strong CIs is also evident. A weak CI can be regarded as a result of relaxation from integrating two CIs, its theoretical coverage probability is at least $1-\beta$ (and usually much greater than $1-\beta$ in practice). A strong CI is expected to achieve target coverage probability; however, the bias term in the strong CI could be difficult to estimate. Nevertheless, both weak and strong CIs could be shown to be asymptotically valid in the following theorem.

Theorem 3.2.5. *Under (Weak) Assumption 3.2.1, the weak CIs defined in (3.2.16) and (3.2.17) are asymptotically valid, i.e.,*

$$\begin{cases} \lim_{N\to\infty}\lim_{M\to\infty} P\{CL^w_{var} \le VaR_\alpha \le CU^w_{var}\} \ge 1-\beta, \\ \lim_{N\to\infty}\lim_{M\to\infty} P\{CL^w_{cvar} \le CVaR_\alpha \le CU^w_{cvar}\} \ge 1-\beta, \end{cases} \tag{3.2.22}$$

where CL^w_{var}, CU^w_{var}, CL^w_{cvar} and CU^w_{cvar} denote the boundaries of weak CIs. Furthermore, under (Strong) Assumption 3.2.2, the strong CIs defined in (3.2.20) and (3.2.21) are asymptotically valid when $N = o_M(M^2)$, i.e.,

$$\begin{cases} \lim_{N,M\to\infty} P\{CL^s_{var} \le VaR_\alpha \le CU^s_{var}\} \ge 1-\beta, \\ \lim_{N,M\to\infty} P\{CL^s_{cvar} \le CVaR_\alpha \le CU^s_{cvar}\} \ge 1-\beta, \end{cases} \tag{3.2.23}$$

when $N = o_M(M^2)$*, where* CL^s_{var}*,* CU^s_{var}*,* CL^s_{cvar} *and* CU^s_{cvar} *denote the boundaries of strong CIs.*

Proof. By Boole's Inequality, to show (3.2.22), it is sufficient to show the following limits

$$\begin{cases} \lim\limits_{N\to\infty}\lim\limits_{M\to\infty} P\{|Err_1| \le 2\frac{t_{1-\beta_I/2,M-1}\widehat{\tau}_{var}}{\sqrt{M}}\} = 1-\beta_I, \\ \lim\limits_{N\to\infty}\lim\limits_{M\to\infty} P\{|Err_2| \le 2\frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{var}}{\sqrt{N}}\} = 1-\beta_O, \\ \lim\limits_{N\to\infty}\lim\limits_{M\to\infty} P\{|Err_3| \le 2\frac{t_{1-\beta_I/2,(1-\alpha)NM-1}\widehat{\tau}_{cvar}}{\sqrt{(1-\alpha)NM}}\} = 1-\beta_I, \\ \lim\limits_{N\to\infty}\lim\limits_{M\to\infty} P\{|Err_4| \le 2\frac{t_{1-\beta_O/2,N-1}\widehat{\sigma}_{cvar}}{\sqrt{N}}\} = 1-\beta_O. \end{cases}$$

where recall Err_1 - Err_4 are defined in (3.2.6) and (3.2.7). In view of the fact that a Student's t-distribution converges to a standard normal distribution as the degree of freedom goes to infinity, the almost sure convergence of variance estimators by SLLN, and the consistency of kernel density estimation, these limits naturally hold following Theorem 3.2.3. Similarly, (3.2.23) can be established. □

3.3 Budget Allocation

In a practical simulation experiment, usually there is a computation budget that affects the sample sizes N and M. Intuitively, the outer sample size N determines the simulation error due to input uncertainty, while the inner sample size M determines the simulation error due to stochastic uncertainty. Therefore, choosing appropriate N and M is critical to balance the trade-off between capturing input uncertainty and capturing stochastic uncertainty, and improve overall experiment efficiency. As shown in previous section, under Strong Assumption, the error of nested risk estimator could be decomposed into two components corresponding to the one layer simulation error caused by input uncertainty and the simulation bias caused by stochastic uncertainty. Within this framework, [40] minimizes the asymptotic MSE (equals variance plus bias square) of nested risk estimator using its asymptotic representation that depends on N and M. The result is an (asymptotically) optimal budget allocation ($N = O_M(M^2)$)

that balances between outer sampling error and inner sampling bias.

In this section, we will consider an alternative approach to improve simulation efficiency under Weak Assumption. Now the total error of nested risk estimators is decomposed into two asymptotically normally distributed components that correspond to the simulation error due to input uncertainty and the simulation error due to stochastic uncertainty, respectively. Thus, the optimal budget allocation is determined by minimizing the final CI half width. Our results could be viewed as a complement of existing methods within the framework of efficient nested risk simulation.

In particular, in view of the weak CIs (3.2.16) and (3.2.17), the CI width minimization problem could be formulated as follows. Let $W_{var}(N, M)$ and $W_{cvar}(N, W)$ be the CI half widths for VaR_α and $CVaR_\alpha$, respectively, i.e.,

$$W_{var}(N, M) := \frac{t_{1-\beta_O/2, N-1}\sigma_{var}}{\sqrt{N}} + \frac{t_{1-\beta_I/2, M-1}\tau_{var}}{\sqrt{M}}, \tag{3.3.1}$$

and

$$W_{cvar}(N, M) := \frac{t_{1-\beta_O/2, N-1}\sigma_{cvar}}{\sqrt{N}} + \frac{t_{1-\beta_I/2, (1-\alpha)NM-1}\tau_{cvar}}{\sqrt{(1-\alpha)NM}}. \tag{3.3.2}$$

They are the objective functions in the budget allocation problem. Note that there are four experiment parameters, i.e., β_O, β_I, N and M to be determined. To reduce the number of decision variables and ease the optimization, we pre-select β_O and β_I (a typical selection is $\beta_O = \beta_I = \beta/2$). Next, let us describe the constraints. Let $\Delta(N, M)$ be the total computation cost $c_1 N + c_2 NM$, where c_1 is the cost for generating one input parameter scenario, and c_2 is the cost for generating one response sample. Of course, there could be other criteria such as computation complexity, and these can be minimized in a similar manner. Let CB be the total computation budget,

and consider the following CI (half) width minimization problem

$$\begin{array}{lll} \min\limits_{N,M} & W_{var}(N,M) & \text{OR} \quad \min\limits_{N,M} \quad W_{cvar}(N,M) \\ s.t. & \Delta(N,M) \le CB & s.t. \quad \Delta(N,M) \le CB \\ & N \ge \Gamma_0,\ M \ge \Gamma_0 & \quad\ N \ge \Gamma_0,\ M \ge \Gamma_0,\ (1-\alpha)NM \ge \Gamma_0 \\ & N, M \in \mathbb{Z}^+ & \quad\ N, M \in \mathbb{Z}^+ \end{array} \tag{3.3.3}$$

The constraints $N \ge \Gamma_0,\ M \ge \Gamma_0$ and $(1-\alpha)NM \ge \Gamma_0$ are imposed to ensure the validity of t-statistics, and a typical choice for Γ_0 is 30.

The practical challenge in solving minimization problem (3.3.3) is the lack of knowledge for the "variance parameters" σ_{var}, τ_{var}, σ_{cvar} and τ_{cvar} in the objective function, since they usually need to be estimated using the samples generated in the simulation experiment. A natural fix is to run a pilot experiment with a small fraction of total budget, obtain crude approximations of these unknown parameters, and solve the problem with these approximate parameter values.

To this end, let us use $\widetilde{\sigma}_{var}$, $\widetilde{\tau}_{var}$, $\widetilde{\sigma}_{cvar}$ and $\widetilde{\tau}_{cvar}$ to denote the crude approximations of σ_{var}, τ_{var}, σ_{cvar} and τ_{cvar} from the pilot run, respectively. Theoretically, they could be the corresponding sample average estimates. However, this method might be extremely inaccurate and unstable because it involves rare-event simulation with few samples. For example, recall that $\sigma_{cvar} = \sqrt{Var\left[(H(\theta) - VaR_\alpha)^+\right]}/(1-\alpha)$, and its square can be rewritten as

$$\begin{aligned} \sigma^2_{cvar} &= \frac{Var\left[(H(\theta) - VaR_\alpha)^+\right]}{(1-\alpha)^2} \\ &= \frac{1}{(1-\alpha)^2}\left\{\mathbb{E}\left[(H(\theta) - VaR_\alpha)^2 I\{H(\theta) \ge VaR_\alpha\}\right] - \left(\mathbb{E}\left[(H(\theta) - VaR_\alpha)^+\right]\right)^2\right\}. \end{aligned} \tag{3.3.4}$$

By definition we have $CVaR_\alpha = VaR_\alpha + \frac{1}{1-\alpha}\mathbb{E}\left[(H(\theta) - VaR_\alpha)^+\right]$. Hence, estimation of σ^2_{cvar} is at least as inefficient as estimation of $CVaR_\alpha$—one of the our initial goals. If we use the naive sample average to estimate the two expectation terms in (3.3.4), most

samples would be ineffective. In fact, theoretically only $(1-\alpha)$ fraction of samples will be effective. Because α is close to 1 (usually $\alpha = 0.95$ or 0.99), the percentage of effective samples is close to 0. To be more specific, suppose $\alpha = 0.99$ and 100 scenarios of mean response ($N = 100$) are generated in the pilot run. Then theoretically only one scenario will be used to estimate σ^2_{cvar} since the rest 99 scenarios result in a simple value of 0. Obviously, the one-scenario estimate is very likely to be far from the true value. Intuitively, the naive sample average method is problematic because the information about the distribution carried by the ineffective samples is not used. In contrast, a good estimation method should try to make use of the information carried by all the samples. For example, using (adaptive) importance sampling will turn most of the ineffective samples into effective samples, and therefore improve the accuracy. This approach is not readily applicable in our model because the lack of knowledge for p.d.f. of the mean response distribution as well as the limitation of budget in learning the distribution in the pilot run.

Next, we will describe a new approach to estimating the variance parameters ($\widetilde{\sigma}_{var}$, $\widetilde{\tau}_{var}$, $\widetilde{\sigma}_{cvar}$ and $\widetilde{\tau}_{cvar}$) that exploits the information carried by all the samples generated in the pilot run. In view of the definitions of variance parameters, i.e.,

$$\sigma^2_{var} = \alpha(1-\alpha)/f^2(VaR_\alpha), \quad \tau^2_{var} = \mathbb{E}[\tau^2_\theta | H(\theta) = VaR_\alpha],$$

and

$$\sigma^2_{cvar} = Var\left[(H(\theta) - VaR_\alpha)^+\right]/(1-\alpha)^2, \quad \tau^2_{cvar} = \mathbb{E}[\tau^2_\theta | H(\theta) \geq VaR_\alpha],$$

the challenges are two folds: (i) the lack of explicit formula for $f(\cdot)$, i.e., p.d.f. of $H(\theta)$; (ii) the lack of functional representation of τ^2 in $H(\theta)$, i.e., $\tau^2(y) \stackrel{\triangle}{=} \mathbb{E}[\tau^2_\theta | H(\theta) = y]$ is not available.

To address the first challenge, we apply the "density projection" technique to project the discrete empirical distribution of $H(\theta)$ onto a parameterized family of

continuous distributions. Specifically, a *projection mapping* from a space of probability distributions $\mathcal{P}$ to another space consisting of a parameterized family of densities $\mathcal{F}$, denoted as $Proj_{\mathcal{F}} : \mathcal{P} \to \mathcal{F}$, is defined by

$$Proj_{\mathcal{F}}(g) \stackrel{\triangle}{=} \arg\min_{f \in \mathcal{F}} D_{KL}(g \parallel f), \quad \forall g \in \mathcal{P}, \tag{3.3.5}$$

where $D_{KL}(g \parallel f)$ denotes the *Kullback-Leibler (K-L) divergence* (or *relative entropy*) between g and f, which is

$$D_{KL}(g \parallel f) \stackrel{\triangle}{=} \int g(x) \log \frac{g(x)}{f(x)} dx. \tag{3.3.6}$$

Here note that the densities g and f are assumed to have the same support. Hence, the projection of g on $\mathcal{F}$ has the minimum K-L divergence from g among all the densities in $\mathcal{F}$. Loosely speaking, the projection of g on $\mathcal{F}$ is the "best" approximation of g one can find in $\mathcal{F}$. When $\mathcal{F}$ is an exponential family of densities, the minimization problem (3.3.5) has an analytical solution and can be carried out easily. The exponential families include many common families of densities, such as Gaussian, Binomial, Poisson and Gamma. By choosing an appropriate exponential family of densities as the target space in projecting the empirical distribution of $H(\theta)$, the resulting density could be regarded as an approximation of $f(\cdot)$ that admits an explicit formula. Therefore, $\widetilde{\sigma}^2_{var}$ and $\widetilde{\sigma}^2_{cvar}$ can be computed by simple numerical calculation instead of sample averaging. More importantly, this technique makes use of information carried by all the samples.

To address the second challenge, we apply regression for $\tau^2(y)$ onto the space of $H(\theta)$ and use samples from the pilot run to train the regression model. Simple numerical tests show that a polynomial regression with basis functions consisting of polynomial (degree≤ 3) functions of $H(\theta)$ is sufficiently good. Finally, $\widetilde{\tau}^2_{var}$ and $\widetilde{\tau}^2_{cvar}$ are computed via numerical integration.

After plugging the approximate variance parameters $\widetilde{\sigma}_{var}$, $\widetilde{\tau}_{var}$, $\widetilde{\sigma}_{cvar}$ and $\widetilde{\tau}_{cvar}$ into

problem (3.3.3), the remaining challenge is how to solve the optimization problem efficiently. Solving it analytically to optimality is unlikely because the objective function does not exhibit structural properties (e.g., convexity). Alternatively, we can enumerate a reasonable amount of candidate allocation schemes (e.g., a two-dimensional grid of feasible allocation schemes), and choose the one scheme that yields the smallest CI width.

We point out that it is worthwhile to consider a more sophisticated budget allocation scheme so that the inner sample size also varies across different input parameter scenarios. For example, in VaR_α estimation, the input parameter scenarios that heavily affect estimation accuracy are the ones with mean responses close to VaR_α. For a specific scenario, it affects the VaR_α estimation if the estimated mean response for the scenario falls into one side of VaR_α while the true mean response for the scenario falls into the other side. Then the inner sample size for this scenario should be increased to reduce the probability of such event. This problem has been studied in the setting of nested credit risk assessment using ranking and selection (see e.g., [18]) and screening (see e.g., [58]), etc.

3.4 Numerical Experiments

We first use the simple numerical example in [40] to compare the CI procedures under Weak Assumption and Strong Assumption (and referred to as weak CI procedure and strong CI procedure), respectively. In particular, consider $H(\theta;\xi) = \mathcal{N}(0,1)+\mathcal{N}(0,1)$, the summation of two independent standard normal random variables. In [40], the first $\mathcal{N}(0,1)$ represents the (outer-layer) portfolio loss distribution and the second $\mathcal{N}(0,1)$ represents the (inner-layer) pricing error. Clearly, this example does not fit into our input uncertainty framework. The reason for using this example is that the exact risk values, and all the variance and bias parameters have closed-form expressions. Thus, the comparison between weak CI procedure and strong CI procedure is

precise. The performance measures of interest include CI width and actual coverage probability, i.e., the probability that the true risk value falls into the simulated CI. In particular, we will run 1000 independent replications to compute the two performance measures, in which the optimal budget allocation scheme via minimizing CI width is employed in weak CI procedure and the optimal budget allocation scheme via minimizing MSE from [40] is employed in strong CI procedure. The results for VaR (results for CVaR are similar, and thus omitted) are summarized in Table 3.4.1.

Table 3.4.1: Comparison of Two 95% CI Procedures—VaR.

	CI under Weak Assumption				**CI under Strong Assumption**			
CB	N_W	M_W	Half CI Width	Cover. Prob.	N_S	M_S	Half CI Width	Cover. Prob.
10^4	212	47	0.65	100%	33	311	0.72	94.7%
10^5	669	149	0.37	100%	70	1446	0.50	95.9%
10^6	2114	473	0.21	100%	149	6716	0.34	95.8%
10^7	6683	1496	0.12	100%	321	31173	0.23	95.8%

$\alpha = 0.95$, $CB = NM + N$—total # samples. (N_W, M_W)—optimal budget allocation by minimizing CI width under weak assumption. (N_S, M_S)—optimal budget allocation by minimizing MSE of risk estimator in [40] under strong assumption. Cover. Prob.—coverage probability, which is obtained via 1000 independent replications. Furthermore, the CI under weak assumption does not take bias into account while the CI under strong assumption does.

The numerical results show that: 1) The optimal budget allocation schemes for weak CI procedure and strong CI procedure could be drastically different. 2) In general, the weak CI procedure yields narrower CIs compared with the strong CI procedure. 3) The weak CIs have coverage probabilities (100%) much greater than the target confidence level (95%) as expected, while the strong CIs have coverage probabilities almost equal to the target confidence level. In this example, the weak CIs appear to be better than the strong CIs because they are narrower and have higher coverage probabilities; however, we should note that the budget allocation scheme for the strong CI procedure aims at minimizing MSE not CI width.

Next, let us consider another example—risk assessment of the $M/M/1$ queueing system in [92] under input uncertainty. In particular, we focus on estimating

the risk of mean sojourn time due to input uncertainty. In the $M/M/1$ queueing system, assume the "true" Poisson customer arrival rate is λ_o, which means the inter-arrival times between customers are independently sampled from an exponential distribution with rate λ_o. Further assume the "true" Exponential service rate is μ_o, which means the service time for each customer is sampled from an exponential distribution with rate μ_o. Here "true" means that the values of λ_o and μ_o are known to us (the judges) but not known to the experimenter. We will mainly follow the experiment parameter set-up in [92], i.e., $\mu_o = 500$ and $\lambda_o = 50, 250, 450$—a range of values corresponding to increasing levels of "true" arrival intensity. To model input uncertainty, we take a Bayesian approach to construct the belief distribution on input parameters—the Poisson arrival rate λ and the exponential service rate μ. Specifically, assume non-informative priors for both λ and μ, i.e., $p(\lambda) \propto 1/\lambda$ and $p(\mu) \propto 1/\mu$. Based on $n = 10, 100, 10000$ historical observations of λ and μ (drawn from the corresponding distributions with "true" parameters), standard Bayesian updating is applied to obtain the posterior distributions of λ and μ. In particular, denote the historical observations of λ by $\mathbf{x} = (x_1, ..., x_n)$. Then sequential Bayesian updating on the posterior distribution of λ is carried out analytically and leads to $p(\lambda|\mathbf{x}) = \lambda^{n-1} \exp(-\lambda \sum_{i=1}^{n} x_i)$, which is a Gamma distribution with shape parameter n and scale parameter $1/(\sum_{i=1}^{n} x_i)$. Similarly, let $\mathbf{y} = (y_1, ..., y_n)$ be the historical observations of μ. Then the posterior distribution of μ is $p(\mu|\mathbf{y}) = \mu^{n-1} \exp(-\mu \sum_{i=1}^{n} y_i)$—a Gamma distribution with shape parameter n and scale parameter $1/(\sum_{i=1}^{n} y_i)$.

The objective is to estimate $VaR_\alpha/CVaR_\alpha$ ($\alpha = 0.90, 0.95, 0.99$) of mean sojourn time w.r.t. the posterior parameter distributions $p(\lambda|\mathbf{x})$ and $p(\mu|\mathbf{y})$, and construct the associated $100(1-\beta)\%$ CIs ($\beta = 0.05$). In particular, we draw $N = 5000$ input parameter scenarios from $p(\lambda|\mathbf{x})$ and $p(\mu|\mathbf{y})$ that satisfies $\lambda < \mu$ (requirement of a stable queue). Furthermore, for each input parameter scenario, we draw $M = 200$

samples of sojourn times by simulating the queue's first 200 sojourn cycles to estimate its mean sojourn time. Finally, VaR_α and $CVaR_\alpha$ of mean sojourn time are estimated via (3.1.4). As for the CI construction, the CI procedure under Weak Assumption is used. The reason for not using the CI procedure under Strong Assumption is that the bias components (see Lemma B.2.2 and B.2.3 in Appendix B.2 for explicit formulas) in the CIs are very difficult to estimate accurately. In fact, our numerical tests show that estimating bias brings new error that overwhelms the bias itself. The simulation results are summarized in Table 3.4.2 and 3.4.3.

Table 3.4.2: VaR (with 95% CI) for the Mean Sojourn Time in an M/M/1 Queue.

λ_o	n	Mean $\mp$ Half CI Width	VaR_{α_1} $\mp$ Half CI Width	VaR_{α_2} $\mp$ Half CI Width	VaR_{α_3} $\mp$ Half CI Width
50	10	$2.4 \times 10^{-3} \mp 3.4 \times 10^{-5}$	$3.7 \times 10^{-3} \mp 7.3 \times 10^{-4}$	$4.5 \times 10^{-3} \mp 1.3 \times 10^{-3}$	$6.5 \times 10^{-3} \mp 1.7 \times 10^{-3}$
50	100	$2.2 \times 10^{-3} \mp 9.7 \times 10^{-6}$	$2.6 \times 10^{-3} \mp 5.1 \times 10^{-4}$	$2.8 \times 10^{-3} \mp 4.8 \times 10^{-4}$	$3.1 \times 10^{-3} \mp 6.1 \times 10^{-4}$
50	10000	$2.2 \times 10^{-3} \mp 6.9 \times 10^{-6}$	$2.4 \times 10^{-3} \mp 5.6 \times 10^{-4}$	$2.5 \times 10^{-3} \mp 4.6 \times 10^{-4}$	$2.7 \times 10^{-3} \mp 5.6 \times 10^{-4}$
250	10	$5.2 \times 10^{-3} \mp 2.1 \times 10^{-4}$	$9.7 \times 10^{-3} \mp 4.4 \times 10^{-3}$	$1.6 \times 10^{-2} \mp 9.5 \times 10^{-3}$	$4.3 \times 10^{-2} \mp 3.1 \times 10^{-2}$
250	100	$4.2 \times 10^{-3} \mp 4.1 \times 10^{-5}$	$5.7 \times 10^{-3} \mp 1.6 \times 10^{-3}$	$6.5 \times 10^{-3} \mp 2.2 \times 10^{-3}$	$8.7 \times 10^{-3} \mp 4.3 \times 10^{-3}$
250	10000	$3.9 \times 10^{-3} \mp 1.8 \times 10^{-5}$	$4.5 \times 10^{-3} \mp 1.1 \times 10^{-3}$	$4.7 \times 10^{-3} \mp 1.7 \times 10^{-3}$	$5.1 \times 10^{-3} \mp 1.4 \times 10^{-3}$
450	10	$9.9 \times 10^{-3} \mp 3.3 \times 10^{-4}$	$2.4 \times 10^{-2} \mp 1.6 \times 10^{-2}$	$3.4 \times 10^{-2} \mp 2.7 \times 10^{-2}$	$5.5 \times 10^{-2} \mp 4.1 \times 10^{-2}$
450	100	$1.8 \times 10^{-2} \mp 3.6 \times 10^{-4}$	$3.5 \times 10^{-2} \mp 2.6 \times 10^{-2}$	$4.2 \times 10^{-2} \mp 2.8 \times 10^{-2}$	$5.3 \times 10^{-2} \mp 3.7 \times 10^{-2}$
450	10000	$2.1 \times 10^{-2} \mp 2.6 \times 10^{-4}$	$3.0 \times 10^{-2} \mp 2.4 \times 10^{-2}$	$3.4 \times 10^{-2} \mp 2.5 \times 10^{-2}$	$4.1 \times 10^{-2} \mp 2.7 \times 10^{-2}$

The experiment parameters are: $\mu_o = 500$, the outer sample size $N = 5000$, the inner sample size $M = 200$, $\alpha_1 = 0.90$, $\alpha_2 = 0.95$, $\alpha_3 = 0.99$.

We have the following observations:

- In general, there are significant gaps between expectations of mean sojourn time (column 3) w.r.t. input uncertainty and VaR/CVaR of mean sojourn time (columns 4 to 6) w.r.t. input uncertainty. It implies that using risk formulation

Table 3.4.3: CVaR (with 95% CI) for the Mean Sojourn Time in an M/M/1 Queue.

λ_o	n	Mean $\mp$ Half CI Width	$CVaR_{\alpha_1}$ $\mp$ Half CI Width	$CVaR_{\alpha_2}$ $\mp$ Half CI Width	$CVaR_{\alpha_3}$ $\mp$ Half CI Width
50	10	$2.4 \times 10^{-3} \mp 3.4 \times 10^{-5}$	$5.0 \times 10^{-3} \mp 2.8 \times 10^{-4}$	$6.0 \times 10^{-3} \mp 4.7 \times 10^{-4}$	$9.0 \times 10^{-3} \mp 1.6 \times 10^{-3}$
50	100	$2.2 \times 10^{-3} \mp 9.7 \times 10^{-6}$	$2.8 \times 10^{-3} \mp 4.9 \times 10^{-5}$	$2.9 \times 10^{-3} \mp 6.9 \times 10^{-5}$	$3.2 \times 10^{-3} \mp 1.6 \times 10^{-4}$
50	10000	$2.2 \times 10^{-3} \mp 6.9 \times 10^{-6}$	$2.6 \times 10^{-3} \mp 3.3 \times 10^{-5}$	$2.6 \times 10^{-3} \mp 4.7 \times 10^{-5}$	$2.8 \times 10^{-3} \mp 9.8 \times 10^{-5}$
250	10	$5.2 \times 10^{-3} \mp 2.1 \times 10^{-4}$	$2.1 \times 10^{-2} \mp 2.4 \times 10^{-3}$	$3.1 \times 10^{-2} \mp 4.2 \times 10^{-3}$	$5.3 \times 10^{-2} \mp 9.6 \times 10^{-3}$
250	100	$4.2 \times 10^{-3} \mp 4.1 \times 10^{-5}$	$7.0 \times 10^{-3} \mp 3.3 \times 10^{-4}$	$7.8 \times 10^{-3} \mp 5.6 \times 10^{-4}$	$1.0 \times 10^{-2} \mp 2.0 \times 10^{-3}$
250	10000	$3.9 \times 10^{-3} \mp 1.8 \times 10^{-5}$	$4.8 \times 10^{-3} \mp 9.7 \times 10^{-5}$	$4.9 \times 10^{-3} \mp 1.4 \times 10^{-4}$	$5.3 \times 10^{-3} \mp 3.2 \times 10^{-4}$
450	10	$9.9 \times 10^{-3} \mp 3.3 \times 10^{-4}$	$3.8 \times 10^{-2} \mp 3.0 \times 10^{-3}$	$4.7 \times 10^{-2} \mp 4.6 \times 10^{-3}$	$6.8 \times 10^{-2} \mp 1.1 \times 10^{-2}$
450	100	$1.8 \times 10^{-2} \mp 3.6 \times 10^{-4}$	$4.3 \times 10^{-2} \mp 2.4 \times 10^{-3}$	$4.9 \times 10^{-2} \mp 3.3 \times 10^{-3}$	$5.8 \times 10^{-2} \mp 7.8 \times 10^{-3}$
450	10000	$2.1 \times 10^{-2} \mp 2.6 \times 10^{-4}$	$3.5 \times 10^{-2} \mp 1.7 \times 10^{-3}$	$3.8 \times 10^{-2} \mp 2.4 \times 10^{-3}$	$4.4 \times 10^{-2} \mp 5.7 \times 10^{-3}$

The experiment parameters are: $\mu_o = 500$, the outer sample size $N = 5000$, the inner sample size $M = 200$, $\alpha_1 = 0.90$, $\alpha_2 = 0.95$, $\alpha_3 = 0.99$.

is quite necessary for accurate risk assessment/control.

- As the number of historical observations increases, VaR/CVaR of mean sojourn time decreases, which indicates that the risk of the system under input uncertainty decreases. Intuitively, as input data size increases, the belief distribution on input parameter becomes more concentrated on the values close to the "true" parameter. Therefore, loosely speaking, the distribution of mean response is also more concentrated on the values close to the "true" mean response, and essentially reduce the risk under input uncertainty.

- As arrival traffic intensifies (λ_o increases) and approaches the service rate μ_o, the system becomes less stable and the risk under input uncertainty is more significant. Therefore, more input data is needed to reduce the risk under input uncertainty to an acceptable level.

We further study the associated budget allocation problem. Note that for VaR estimation and CVaR estimation, the budget allocation problem might yield different optimal allocation schemes. In particular, let $\Delta(N, M) = NM + N$ and $CB = 5 \times 10^5$. $N_{pilot} = 50$ outer scenarios and $M_{pilot} = 100$ inner samples for each scenario are used in the pilot run to guide the budget allocation in the actual run. In total, only 1% percent of total budget is consumed, and the budget for the actual run is barely affected. To exhibit the effectiveness of pilot run, we plot the CI widths for different N in Figure 3.4.1, where the blue curves are the CI widths calculated using variance parameters estimated via density projection and regression in the pilot run, and the red curves are the CI widths calculated using the "true" variance parameters obtained by "simulation-to-death" (i.e., running an extremely large number of simulation replications to have a very accurate estimate, which can be regarded as the true parameter values).

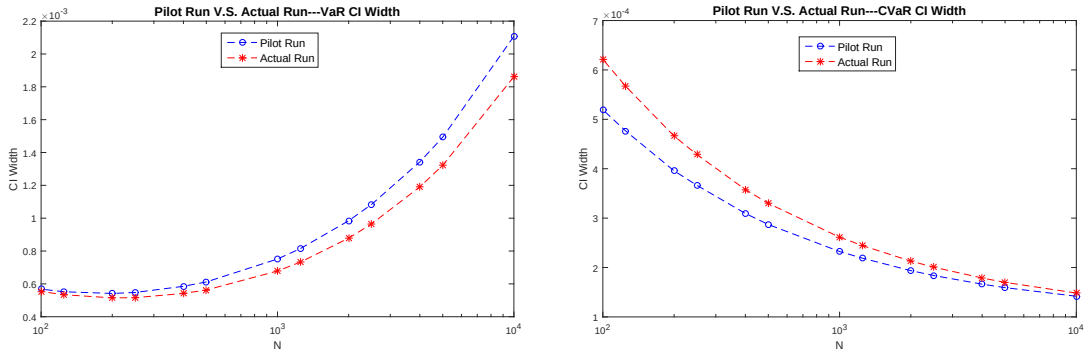

Figure 3.4.1: VaR/CVaR CI Width: Pilot Run V.S. Actual Run.
The experiment parameters are: $\lambda_o = 150$, $\mu_o = 500$, $\alpha = 0.95$, and the size of input data $n = 10$.

- In both plots, although there is a non-negligible gap between the CI width based on the parameters estimated from the pilot run and the true CI width, the curves follow the same trend and their minima coincide, which implies that solving the budget allocation problem would be able to identify the optimal budget allocation scheme. In light of the fact that only 1% of the total computation budget is used, this indicates that the budget allocation problem and

its solution strategy provide extremely effective guidance in determining good budget allocation schemes.

- By comparing the maxima and minima of blue/red curves, we can see that in general the CIs constructed with the best budget allocation schemes are 3 to 4 times narrower than the CIs constructed with the worst budget allocation schemes. When the total budget is very limited, solving the budget allocation problem is very beneficial.

- The best budget allocation schemes for VaR estimation and CVaR estimation are drastically different, where the optimal outer sample size N for minimizing VaR CI width is close to 10^2 and the optimal N for minimizing CVaR CI width is close to 10^4. It is very different from the result by [40] that the optimal N and M for minimizing MSE of nested VaR and CVaR estimators are both $N = O(M^2)$. This is because the budget allocation problem in our work and the budget allocation problem in [40] have different objective criteria, and lead to different optimal solutions.

- Another phenomenon worth mentioning is that CVaR CI width appears to be decreasing in N (see right half of Figure 3.4.1). This is a result of objective function (3.3.2) and budget constraint's special structures. It is easy to see that, as N increases, the first term in (3.3.2) decreases but the second term remains almost unchanged since $NM \approx N + NM = CB$. Therefore, the optimal solution is to increase the outer sample size N and decrease the inner sample size M as much as possible, until M hits the low bound Γ_0.

In conclusion, the simulation results for the $M/M/1$ queueing system provide empirical evidences for the importance and necessity of risk assessment/control of mean response w.r.t. input uncertainty in stochastic systems, as well as the advantages of solving the associated budget allocation problem for efficient nested simulation.

3.5 Conclusion

In the present chapter, we introduce risk measures in stochastic simulation under input certainty, which provide rigorous quantifications of the extreme behavior of mean response in all possible input models. In particular, we use nested Monte Carlo simulation to estimate the VaR/CVaR of mean response w.r.t. input uncertainty. We prove the asymptotical properties (consistency and normality) of nested risk estimators in different limiting senses under different sets of regularity conditions. We further use those properties to construct (asymptotically valid) CIs, and propose a practical framework of optimal budget allocation for improving the efficiency of nested risk simulation. The work in this chapter can be viewed as a starting point of research on more general risk measures for risk assessment in stochastic simulation under input uncertainty.

On the other hand, the naive estimators considered here could be restrictive in risk assessment for large-scale systems under input uncertainty, because of the inefficiency of naive rare-event simulation. The budget allocation problem solved in this chapter partially addresses this issue in the sense that it leads to good outer versus inner sample size tradeoff in reducing CI width. Developing more sophisticated budget allocation schemes will be a promising direction of future research.

CHAPTER IV

OPTIMIZATION OF RISK MEASURES VIA GRADIENT-BASED STOCHASTIC ADAPTIVE SEARCH

In this chapter, we will present the framework of risk optimization via Gradient-based Stochastic Adaptive Search (GASS). This work is also summarized in [88].

The rest of the chapter is organized as follows. In Section 4.1, we will formulate the CVaR optimization problem. Then we extend GASS algorithm, which is originally developed for deterministic non-differentiable optimization problems, to the CVaR optimization problem. The detailed algorithms are presented in Section 4.2, in which Algorithm 4.2.1 (referred to as "GASS-CVaR") is a straightforward extension of GASS and Algorithm 4.2.2 (referred to as "GASS-CVaR-ARL") further incorporates an updating rule for adaptive adjustments of the risk level. Convergence analysis of both algorithms are presented in Section 4.3. In Section 4.4, we illustrate the performance of the proposed algorithms by carrying out numerical tests on several benchmark loss functions. We conclude the chapter in Section 4.5.

4.1 General Framework

Consider a scalar loss function of the form $l(x, \xi_x)$, where $x \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ represents the decision variables, and ξ_x represents the randomness in the loss function and its distribution may or may not depend on x. The loss function $l(x, \xi_x)$ can be evaluated either directly or through simulation. Furthermore, to ease the presentation, assume $l(x, \xi_x)$ admits an almost everywhere (a.e.) positive and continuous probability density function (p.d.f.) $p(t; x)$, and thus a continuous and strictly increasing cumulative distribution function (c.d.f.) $P(t; x)$ for all $x \in \mathcal{X}$. The objective is to minimize the

CVaR of the loss function $l(x;\xi_x)$ at a risk level of interest α^* $(0 < \alpha^* < 1)$ with respect to (w.r.t.) $x \in \mathcal{X}$. That is, to solve the following stochastic optimization problem:

$$\min_{x\in\mathcal{X}} C_{\alpha^*}(x) \triangleq CVaR_{\alpha^*}\left(l(x,\xi_x)\right), \quad \text{or equivalently,} \quad \max_{x\in\mathcal{X}} -C_{\alpha^*}(x), \tag{4.1.1}$$

where $CVaR_{\alpha^*}\left(l(x,\xi_x)\right)$ is defined by

$$\begin{aligned} CVaR_{\alpha^*}\left(l(x,\xi_x)\right) &\triangleq \mathbb{E}_{\xi_x}\left[l(x,\xi_x)|l(x,\xi_x) \geq V_{\alpha^*}(x)\right] \\ &= \frac{1}{1-\alpha^*}\mathbb{E}_{\xi_x}\left[l(x,\xi_x)\mathbb{1}\left\{l(x,\xi_x) \geq V_{\alpha^*}(x)\right\}\right] \\ &= V_{\alpha^*}(x) + \frac{1}{1-\alpha^*}\mathbb{E}_{\xi_x}\left[\left(l(x,\xi_x) - V_{\alpha^*}(x)\right)^+\right], \end{aligned}$$

where $\mathbb{1}\{A\}$ is 1 if event A is true and 0 otherwise, $(u)^+ = \max(u,0)$, and $V_{\alpha^*}(x)$ is the VaR of $l(x,\xi_x)$ at risk level α^*, i.e.,

$$V_{\alpha^*}(x) = VaR_{\alpha^*}\left(l(x,\xi_x)\right) \triangleq \inf\{t : P(t;x) \geq \alpha^*\} = P^{-1}(\alpha^*;x).$$

Note that the inverse c.d.f. $P^{-1}(\alpha^*;x)$ exists because $P(t;x)$ is strictly increasing in t. We also follow the standard assumption that $C_{\alpha^*}(x)$ is bounded from below and above on $\mathcal{X}$, i.e., $\exists C_{lb} > -\infty, C_{ub} < \infty$ s.t. $C_{lb} < C_{\alpha^*}(x) < C_{ub}$, $\forall x \in \mathcal{X}$.

Problem (4.1.1) might be difficult to solve when $l(x;\xi_x)$ lacks structural properties such as convexity and differentiability. Most of the gradient-based algorithms might fail. Instead, we seek model-based methods to solve problem (4.1.1). In principle, we could extend GASS algorithm in [87] to solve it.

4.1.1 Main Idea

Similar to many other model-based methods, the main idea of GASS is to introduce a parameterized sampling distribution over the solution space, and update the parameters of the sampling distribution iteratively towards the promising region of the solution space. Let us illustrate the main idea in a general framework, where one aims to maximize a deterministic function $L(x)$ over $x \in \mathcal{X}$.

We introduce a parameterized family of densities $\{f(x;\theta) : \theta \in \Theta \subset \mathbb{R}^{d_\theta}\}$ as the sampling distribution, where θ represents the parameter that will be updated over iterations. Consider a simple reformulation as follows:

$$H(\theta) \stackrel{\triangle}{=} \int L(x) f(x;\theta) dx.$$

Then $H(\theta) \leq L(x^*) = L^*$, where x^* denotes the optimal solution or one of the optima, and L^* denotes the optimal function value on $\mathcal{X}$. Note that the equality is achieved if and only if all the probability mass of $f(x;\theta)$ concentrates on a subset of the set of global optima. Given the existence of such a θ, we can solve the reformulated problem $\max_{\theta\in\Theta} H(\theta)$ instead of the original problem, since the optimal parameter will recover the optimal solution and the optimal function value.

The advantage of the reformulated problem over the original problem is that it is differentiable in θ under mild regularity conditions on $f(x;\theta)$, and the gradient is easy to derive as follows:

$$\begin{aligned}\bigtriangledown_\theta H(\theta) &= \bigtriangledown_\theta \int L(x) f(x;\theta) dx = \int L(x) \frac{\bigtriangledown_\theta f(x;\theta)}{f(x;\theta)} f(x;\theta) dx \\ &= \mathbb{E}_{f(\cdot;\theta)} \left[L(x) \bigtriangledown_\theta \ln f(x;\theta) \right].\end{aligned}$$

Note that an unbiased estimator of $\bigtriangledown_\theta H(\theta)$ could be obtained by drawing samples $x^i \stackrel{i.i.d.}{\sim} f(x;\theta), i = 1, ..., N$, evaluating $L(x^i) \bigtriangledown_\theta \ln f(x^i;\theta)$, and taking the sample average of $\{L(x^i) \bigtriangledown_\theta \ln f(x^i;\theta) : i = 1, ..., N\}$. Therefore, one could solve the reformulated problem via a (stochastic) gradient-based method. Specifically, the method iteratively carries out the following two steps:

1. Generate candidate solutions according to the sampling distribution.

2. Based on the evaluation of the candidate solutions, update the parameter of the sampling distribution via gradient search.

Intuitively, it combines the relative fast convergence of gradient search with the robustness of model-based optimization in terms of maintaining a global exploration of the solution space.

4.1.2 Review of GASS

Based on the above main idea, now let us describe the full-blown GASS algorithm. We introduce a shape function $S_\theta : \mathbb{R} \to \mathbb{R}^+$, where the subscript θ signifies the possible dependence of the shape function on the parameter θ, and it satisfies the following conditions: for every θ, $S_\theta(y)$ is strictly increasing in y and bounded from below and above for finite y; moreover, for every fixed y, $S_\theta(y)$ is continuous in θ. The purpose is to make the objective function positive while preserving the order of the solutions and in particular the optimal solution. Moreover, the shape function adds flexibility to the algorithm by giving a user the freedom to choose a weighting scheme on the samples based on sample function evaluations. For example, a good choice of the shape function $S_\theta(\cdot)$ is

$$S_\theta(L(x)) = \frac{1}{1 + \exp(-S_0(L(x) - \gamma_\theta))}, \tag{4.1.2}$$

where S_0 is a large positive constant, and γ_θ is the $(1-\rho)$-quantile

$$\gamma_\theta \stackrel{\triangle}{=} \sup_r \left\{ r : P_{f(\cdot;\theta)} \left\{ x \in \mathcal{X} : L(x) \geq r \right\} \geq \rho \right\}, \tag{4.1.3}$$

where $P_{f(\cdot;\theta)}\{A\}$ denotes the probability of event A w.r.t. $f(\cdot;\theta)$. Notice that $S_\theta(\cdot)$ could be viewed as a continuous approximation of the indicator function $\mathbb{1}\{L(x) \geq \gamma_\theta\}$ that eliminates the $L(x)$ values below γ_θ.

For an arbitrary but fixed $\theta' \in \Theta$, let us define

$$H(\theta;\theta') \stackrel{\triangle}{=} \int S_{\theta'}(L(x)) f(x;\theta) dx, \quad \text{and} \quad h(\theta;\theta') \stackrel{\triangle}{=} \ln H(\theta;\theta'). \tag{4.1.4}$$

By the condition on the shape function and the fact that $\ln(\cdot)$ is a strictly increasing function, solving the original problem is equivalent to solving the problem $\max_{\theta\in\Theta} h(\theta;\theta')$ for any fixed θ'. Following the main idea outlined before, [87] propose a stochastic search algorithm that iteratively carries out the following two steps:

1. Generate candidate solutions from $f(x;\theta_k)$, where θ_k is the parameter obtained at iteration k.

2. Update the parameter θ_k using a Newton-like iteration for $\max_\theta h(\theta;\theta_k)$.

Note that the second step requires to compute the gradient and Hessian of $h(\theta;\theta_k)$, which, as shown by [87], have analytical expressions as the expectations under certain probability measures. In particular, if the sampling distributions are chosen to be an exponential family of densities in the following Definition 4.1.1, then these expressions can be further simplified.

Definition 4.1.1. *A family $\{f(x;\theta):\theta\in\Theta\}$ is an exponential family of densities if it satisfies*

$$f(x;\theta)=\exp\left\{\theta^T\Gamma(x)-\eta(\theta)\right\},$$

where $\Gamma(x)=[\Gamma_1(x),...,\Gamma_d(x)]^T$ is the vector of sufficient statistics,

$$\eta(\theta)=\ln\{\int\exp(\theta^T\Gamma(x))dx\}$$

is the normalization factor to ensure $f(x;\theta)$ is a p.d.f., and $\Theta=\{\theta:|\eta(\theta)|<\infty\}$ is the natural parameter space with a nonempty interior.

Proposition 4.1.1 below provides the corresponding analytical expressions of the gradient and Hessian of $h(\theta;\theta')$ when an exponential family of densities is used as the sampling distribution. We refer to [87] for the detailed derivations.

Proposition 4.1.1. *If $\{f(x;\theta):\theta\in\Theta\}$ is an exponential family of densities, then the gradient and Hessian of $h(\theta;\theta')$ in (4.1.4) have closed-form expressions as follows:*

$$\begin{cases}\bigtriangledown_\theta h(\theta;\theta')\big|_{\theta=\theta'}=\mathbb{E}_{q(\cdot;\theta')}\left[\Gamma(x)\right]-\mathbb{E}_{\theta'}\left[\Gamma(x)\right],\\ \bigtriangledown^2_\theta h(\theta;\theta')\big|_{\theta=\theta'}=Var_{q(\cdot;\theta')}\left[\Gamma(x)\right]-Var_{\theta'}\left[\Gamma(x)\right],\end{cases}\tag{4.1.5}$$

where

$$q(x;\theta')=\frac{S_{\theta'}(L(x))f(x;\theta')}{\int S_{\theta'}(L(x))f(x;\theta')dx}$$

is a "re-weighted" p.d.f., $\mathbb{E}_{q(\cdot;\theta')}[\cdot]$ and $Var_{q(\cdot;\theta')}[\cdot]$ denote the expectation and variance w.r.t. $q(\cdot;\theta')$, respectively; $\mathbb{E}_{\theta'}[\cdot]$ and $Var_{\theta'}[\cdot]$ denote the expectation and variance w.r.t. $f(\cdot;\theta')$, respectively.

Note that the Hessian $\bigtriangledown_\theta^2 h(\theta;\theta')\big|_{\theta=\theta'}$ might not be negative semi-definite. To ensure the parameter updating is along the ascent direction of $h(\theta;\theta')$ in a Newton-like scheme, we approximate $\bigtriangledown_\theta^2 h(\theta;\theta')\big|_{\theta=\theta'}$ by a negative-definite term $-(Var_{\theta'}[T(x)] + \epsilon I)$, which is a slight perturbation of the second term in $\bigtriangledown_\theta^2 h(\theta;\theta')\big|_{\theta=\theta'}$. Here ϵ is a small positive number and I is an identity matrix of proper dimension. Then, a Newton-like updating of θ is as follows:

$$\begin{aligned}\theta_{k+1} &= \Pi_\Theta \left\{ \theta_k + \beta_k \left(Var_{\theta_k}[\Gamma(x)] + \epsilon I \right)^{-1} \bigtriangledown_\theta h(\theta;\theta_k)\big|_{\theta=\theta_k} \right\} \\ &= \Pi_\Theta \left\{ \theta_k + \beta_k \left(Var_{\theta_k}[\Gamma(x)] + \epsilon I \right)^{-1} \left(\mathbb{E}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)] \right) \right\}, \quad (4.1.6)\end{aligned}$$

where β_k is a positive step-size, $\mathbb{E}_{q_k}[\cdot]$ denotes the expectation w.r.t. $q(\cdot;\theta_k)$, and $\Pi_\Theta\{\cdot\}$ denotes the projection operator that projects an iterate back onto the parameter space Θ by choosing the closest point in Θ.

In practical implementation, we still need to evaluate or estimate the expectation and variance terms in (4.1.6). Notice that the expectation term $\mathbb{E}_{\theta_k}[\Gamma(x)]$ can be calculated analytically in most cases. For example, if the chosen exponential family of densities is the Gaussian family, then $\mathbb{E}_{\theta_k}[\Gamma(x)]$ reduces to the mean and second moment of the Gaussian distribution. The variance term $Var_{\theta_k}[\Gamma(x)]$ might not be directly available, but it could be estimated by the sample variance using the candidate solutions drawn from $f(\cdot;\theta_k)$. Specifically, suppose N_k i.i.d. samples $\{x_k^i : i = 1, ..., N_k\}$ are drawn from $f(x;\theta_k)$, then

$$\widehat{Var}_{\theta_k}[\Gamma(x)] \triangleq \frac{1}{N_k - 1} \sum_{i=1}^{N_k} \Gamma(x_k^i)\Gamma(x_k^i)^T - \frac{1}{N_k^2 - N_k} \left(\sum_{i=1}^{N_k} \Gamma(x_k^i) \right) \left(\sum_{i=1}^{N_k} \Gamma(x_k^i) \right)^T \quad (4.1.7)$$

is the sample estimate of $Var_{\theta_k}[\Gamma(x)]$. The remaining term $\mathbb{E}_{q_k}[\Gamma(x)]$ can be estimated based on the principle of importance sampling, noting that $\mathbb{E}_{q_k}[\Gamma(x)] \propto \int S_{\theta_k}(L(x))\Gamma(x) f(x;\theta_k) dx$. Therefore, we could estimate it by

$$\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] \triangleq \sum_{i=1}^{N_k} w_k^i \Gamma(x_k^i), \quad (4.1.8)$$

where $\{w_k^i : i = 1, ..., N_k\}$ are normalized weights given by

$$w_k^i = \frac{S_{\theta_k}(L(x_k^i))}{\sum_{j=1}^{N_k} S_{\theta_k}(L(x_k^j))}, \quad i = 1, ..., N_k.$$

When the shape function $S_{\theta_k}(\cdot)$ takes a form such as (4.1.2), it has to be estimated by samples as well since $(1-\rho)$-quantile γ_{θ_k} defined in (4.1.3) needs to be estimated by the sample quantile. In this case, let us denote the sample $(1-\rho)$-quantile by $\widehat{\gamma}_{\theta_k}$ and the approximate shape function by $\widehat{S}_{\theta_k}(\cdot)$, respectively. Then, the normalized weights are evaluated according to

$$\widehat{w}_k^i = \frac{\widehat{S}_{\theta_k}(L(x_k^i))}{\sum_{j=1}^{N_k} \widehat{S}_{\theta_k}(L(x_k^j))}, \quad i = 1, ..., N_k,$$

and the term $\mathbb{E}_{q_k}[\Gamma(x)]$ is approximated by

$$\widehat{\mathbb{E}}_{q_k}[\Gamma(x)] \triangleq \sum_{i=1}^{N_k} \widehat{w}_k^i \Gamma(x_k^i). \tag{4.1.9}$$

Eventually, the gradient $g_k := \mathbb{E}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)]$ is approximated by

$$\widehat{g}_k \triangleq \widehat{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)].$$

4.1.3 Extension of GASS to Optimization of CVaR

When the CVaR of the loss function $C_{\alpha^*}(x)$ could be evaluated exactly for all $x \in \mathcal{X}$, we can directly extend the scheme described above to the CVaR minimization problem (4.1.1). Since the loss function is usually evaluated via simulation, its p.d.f. and c.d.f. are generally not available. Thus, $C_{\alpha^*}(x)$ could not be evaluated analytically; however, it could be estimated via Monte Carlo simulation. In particular, suppose M samples of the loss function $\{l(x, \xi_x^1), l(x, \xi_x^2), ..., l(x, \xi_x^M)\}$ are simulated, and then sorted in ascending order as $l(x, \xi_x^{(1)}) \leq l(x, \xi_x^{(2)}) \leq ... \leq l(x, \xi_x^{(M)})$, which forms an empirical loss distribution. A natural estimator of $C_{\alpha^*}(x)$ is the CVaR of the empirical loss distribution, which is defined as follows.

$$\widehat{C}_{\alpha^*}(x) \triangleq \widehat{V}_{\alpha^*}(x) + \frac{1}{M(1-\alpha^*)} \sum_{m=1}^{M} \left(l(x, \xi_x^m) - \widehat{V}_{\alpha^*}(x) \right)^+, \tag{4.1.10}$$

where

$$\widehat{V}_{\alpha^*}(x) \triangleq l\left(x, \xi_x^{(\lceil \alpha^* M \rceil)}\right) \tag{4.1.11}$$

is the VaR of the empirical loss distribution that plays the role of VaR estimator, and $\lceil \alpha^* M \rceil$ is the smallest integer that is greater than or equal to $\alpha^* M$.

Although the estimator $\widehat{C}_{\alpha^*}(x)$ is biased, it is strongly consistent and asymptotic normally distributed under mild regularity assumptions on the distribution of $l(x, \xi_x)$ (see, e.g., Lemma B.1.1 in Appendix B). In principle, we can use it as a replacement for $C_{\alpha^*}(x)$ and plug it into GASS algorithm.

4.2 Algorithms: GASS-CVAR and GASS-CVAR-ARL

Now let us formally present the following algorithm, which is referred to as GASS-CVaR, for simulation optimization of CVaR.

In the initialization step (step 1) of GASS-CVaR (Algorithm 4.2.1), the conditions on the sample size and step size sequences are imposed to facilitate the convergence of the algorithm. They are typical requirements for a stochastic approximation algorithm. In the sampling step (step 2), notice that the CVaR values are estimated. Therefore, the convergence of the original GASS algorithm, which is designed for deterministic optimization, does not directly apply to GASS-CVaR. In the estimating step (step 3), as mentioned before, one common choice of the shape function $S_\theta(\cdot)$ is in the form of (4.1.2). Moreover, the quantile level ρ in (4.1.3) controls the percentile of elite samples that are used to update the sampling distribution at the next iteration, and balances between the exploitation of the neighborhood of the current best solutions and the exploration of the entire solution space. For example, when a smaller ρ is used, less elite samples are used, and thus less emphasis is put on exploration. In the updating step (step 4), the iterate is projected onto a convex and compact subset $\widetilde{\Theta} \subseteq \Theta$ instead of Θ, in order to guarantee numerical stability and fast computation of the projection. In the stopping step (step 5), a common stopping criterion used in

Algorithm 4.2.1 **G**radient-based **A**daptive **S**tochastic **S**earch for Optimization of **CVaR**

1. **Initialization**: Choose an exponential family of densities $\{f(x;\theta) : \theta \in \Theta\}$, and specify a small positive constant ϵ, initial parameter θ_0, sample size sequence $\{N_k\}$ that satisfies $N_k \to \infty$, simulation budget sequence $\{M_k\}$ that satisfies $M_k \to \infty$, and step size sequence $\{\beta_k\}$ that satisfies $\sum_{k=0}^{\infty} \beta_k = \infty, \sum_{k=0}^{\infty} \beta_k^2 < \infty$. Set $k = 0$.

2.**Sampling**: Draw candidate solutions $\{x_k^i \overset{i.i.d.}{\sim} f(x;\theta_k) : i = 1, 2, ..., N_k\}$. For each x_k^i, simulate the loss function scenarios $\{l(x_k^i, \xi_k^{i,j}) : j = 1, ..., M_k\}$, and sort them in ascending order, denoted by

$$l\left(x_k^i, \xi_k^{i,(1)}\right) \leq l\left(x_k^i, \xi_k^{i,(2)}\right) \leq \cdots \leq l\left(x_k^i, \xi_k^{i,(M_k)}\right).$$

Estimate the CVaR of the loss for each candidate solution at target risk level α^*:

$$\widehat{C}_{\alpha^*}(x_k^i) = l\left(x_k^i, \xi_k^{i,(\lceil \alpha^* M_k \rceil)}\right) + \frac{1}{M_k(1-\alpha^*)} \sum_{j=1}^{M_k} \left(l\left(x_k^i, \xi_k^{i,j}\right) - l\left(x_k^i, \xi_k^{i,(\lceil \alpha^* M_k \rceil)}\right)\right)^+.$$

3. **Estimation**: Compute the normalized weights $\widehat{w}_k^i$ according to

$$\widehat{w}_k^i = \frac{\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^i)\right)}{\sum_{j=1}^{N_k} \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^j)\right)}, \quad i = 1, ..., N_k,$$

and then $\mathbb{E}_{q_k}[\Gamma(x)]$ and $Var_{\theta_k}[\Gamma(x)]$ are estimated via

$$\begin{cases} \widehat{\mathbb{E}}_{q_k}[\Gamma(x)] = \sum_{i=1}^{N_k} \widehat{w}_k^i \Gamma(x_k^i), \\ \widehat{Var}_{\theta_k}[\Gamma(x)] = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} \Gamma(x_k^i)\Gamma(x_k^i)^T - \frac{1}{N_k^2 - N_k} \left(\sum_{i=1}^{N_k} \Gamma(x_k^i)\right)\left(\sum_{i=1}^{N_k} \Gamma(x_k^i)\right)^T. \end{cases}$$

Estimate the gradient g_k via

$$\widehat{g}_k = \widehat{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)].$$

4. **Updating**: Update the sampling distribution parameter θ according to

$$\theta_{k+1} = \Pi_{\widetilde{\Theta}}\left\{\theta_k + \beta_k \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)^{-1} \widehat{g}_k\right\},$$

where $\widetilde{\Theta} \subseteq \Theta$ is a non-empty compact and convex constraint set.

5. **Stopping**: Check if some stopping criterion is satisfied. If yes, stop and return the current best sampled solution; else, set $k := k + 1$ and go back to step 2.

practice is that the norm of the gradient falls below a pre-specified threshold.

4.2.1 GASS with Adaptive Risk Level

When the risk level of interest α^* is close to 1 (e.g. $\alpha^* = 0.99$), implementing GASS-CVaR could be computationally expensive, since the CVaR evaluation in step 2 requires a large simulation budget M_k to obtain good CVaR estimators. This issue is more severe as α^* gets closer to 1. For example, for a fixed x, suppose we want to estimate $C_\alpha(x)$ at three different risk levels: $\alpha_1 = 0$, $\alpha_2 = 0.90$, and $\alpha_3 = 0.99$, where note that $C_{\alpha_1=0}(x) = \mathbb{E}[l(x, \xi_x)]$ is the expected loss. To achieve the same accuracy in estimation of CVaR, the corresponding simulation budgets M_1, M_2, and M_3 should result in equal "effective" simulation budgets $(1-\alpha_i)M_i, i = 1, 2, 3$. Therefore, $M_2 = (1-\alpha_1)/(1-\alpha_2) \cdot M_1 = 10 \cdot M_1$ and $M_3 = (1-\alpha_1)/(1-\alpha_3) \cdot M_1 = 100 \cdot M_1$. This implies that the simulation budget required for CVaR estimation could be easily up to tens of times even hundreds of times compared with the simulation budget required for the estimation of expectation.

To save simulation budget and improve the overall efficiency of GASS-CVaR, we propose to initialize the algorithm at a small risk level α_0 (e.g., $\alpha_0 = 0$), and adaptively increase/update the risk level α_k at every iteration such that the target risk level α^* is achieved and the algorithm converges simultaneously. Since a lower risk level implies that a smaller simulation budget M_k is required to achieve desired accuracy for CVaR estimation, the hope is to adaptively save simulation budget at each iteration by solving a problem that is similar to the original one but less computationally expensive. A good updating rule on the risk level should 1) achieve significant budget savings when the algorithm is in the "warm-up" phase, i.e., when it puts more emphasis on exploration of the entire solution space; 2) solve problems that are close to the original one when the algorithm is in the "convergence" phase, i.e., when it puts more emphasis on the exploitation of the promising region that has

been identified, so that good solutions of the original problem can be found. The key to such an updating rule lies in finding an empirical indication on the algorithm's emphasis between the exploration and the exploitation.

Note that GASS-CVaR maintains the structure of a gradient-based optimization scheme, and thus the gradient (even the Hessian) used in the updating step (step 4) could be regarded as the empirical indication on the balance between the exploration and the exploitation. Loosely speaking, when the norm of the gradient is relatively large, the sampling distribution parameter at next iteration θ_{k+1} will differ from θ_k significantly. This means the algorithm is in the "warm-up" phase, where different regions of the solution space are being explored. When the norm of gradient is small, θ_{k+1} is expected to be close to θ_k. This means the algorithm is in the "convergence" phase, where an identified promising region is being exploited. Therefore, it is natural to design the updating rule on the risk level using the information contained in the gradient obtained at every iteration. For example, note that GASS-CVaR converges when the norm of the gradient hits zero. Then one could increase the risk level at every iteration proportionally to the decrease in the norm of the gradient from the previous iteration. In particular, we propose the following version of GASS algorithm with adaptive risk levels, which is referred to as GASS-CVaR-ARL. We do point out that more sophisticated updating rules on the risk level could be incorporated in the future.

In the sampling step (step 2) of GASS-CVaR-ARL (Algorithm 4.2.2), since the current risk level α_k is smaller than the target risk level α^*, we could use a simulation budget M_k that is smaller than the one used in GASS-CVaR to estimate those CVaR values at risk level α_k. For example, suppose one wants to keep the "effective" simulation budget $(1-\alpha_k)M_k$ as a constant. Then, in the initial iterations of the algorithm the budget savings can be up to tens of times even hundreds of times (equal to $(1-\alpha_k)/(1-\alpha^*)$ precisely) since α_k is close to $\alpha_0 = 0$. The sampled best solution

Algorithm 4.2.2 GASS-CVaR with **A**daptive **R**isk **L**evels

1. **Initialization**: Initialize the algorithm similar as in step 1 of GASS-CVaR. Set initial risk level α_0.

2. **Sampling**: Draw candidate solutions and simulate the loss function scenarios as in step 2 of GASS-CVaR. Estimate the CVaR of the loss for each candidate solution at the risk level α_k:

$$\widehat{C}_{\alpha_k}(x_k^i) = l\left(x_k^i, \xi_k^{i,(\lceil \alpha_k M_k \rceil)}\right) + \frac{1}{M_k(1-\alpha_k)} \sum_{j=1}^{M_k} \left(l\left(x_k^i, \xi_k^{i,j}\right) - l\left(x_k^i, \xi_k^{i,(\lceil \alpha_k M_k \rceil)}\right)\right)^+ .$$

Record the best solution x_k^* found at this iteration: $x_k^* = \arg\min_i \ \widehat{C}_{\alpha_k}(x_k^i)$.

3. **Estimation**: Compute the normalized weights $\overline{w}_k^i$ according to

$$\overline{w}_k^i = \frac{\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha_k}(x_k^i)\right)}{\sum_{j=1}^{N_k} \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha_k}(x_k^j)\right)}, \quad i = 1, ..., N_k,$$

and then $\mathbb{E}_{q_k}[\Gamma(x)]$ and $Var_{\theta_k}[\Gamma(x)]$ are estimated via

$$\begin{cases} \overline{\mathbb{E}}_{q_k}[\Gamma(x)] = \sum_{i=1}^{N_k} \overline{w}_k^i \Gamma(x_k^i), \\ \widehat{Var}_{\theta_k}[\Gamma(x)] = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} \Gamma(x_k^i)\Gamma(x_k^i)^T - \frac{1}{N_k^2 - N_k} \left(\sum_{i=1}^{N_k} \Gamma(x_k^i)\right)\left(\sum_{i=1}^{N_k} \Gamma(x_k^i)\right)^T . \end{cases}$$

Estimate the gradient g_k via

$$\bar{g}_k = \overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)] .$$

4. **Updating**: Update the sampling distribution parameter θ according to

$$\theta_{k+1} = \Pi_{\widetilde{\Theta}} \left\{ \theta_k + \beta_k \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I \right)^{-1} \bar{g}_k \right\},$$

where $\widetilde{\Theta} \subseteq \Theta$ is a non-empty compact and convex constraint set; update the risk level according to

$$\alpha_{k+1} = \begin{cases} \alpha^* - \frac{\|\bar{g}_k\|_2}{\|\bar{g}_{k-1}\|_2} (\alpha^* - \alpha_k), & \text{if } \|\bar{g}_k\|_2 < \|\bar{g}_{k-1}\|_2 , \\ \alpha_k, & o/w, \end{cases} \tag{4.2.1}$$

where $\|\cdot\|_2$ is the vector Euclidean norm.

5. **Stopping**: Check if some stopping criterion is satisfied. If yes, stop and return $x^* = \arg\min_k \ \widehat{C}_{\alpha^*}(x_k^*)$ and $\widehat{C}_{\alpha^*}(x^*)$ via simulation; else, set $k := k + 1$ and go back to step 2.

to the CVaR optimization problem at risk level α_k is also recorded. It can be viewed as a good solution to a CVaR optimization problem that is similar to the original one in structure.

In the updating step (step 4), the updating rule (4.2.1) ensures that α_k is non-decreasing, with the hope that α_k will eventually converge to the target risk level α^*. Furthermore, when $\|\bar{g}_k\|_2 < \|\bar{g}_{k-1}\|_2$, we can rewrite (4.2.1) as $\frac{\alpha^*-\alpha_{k+1}}{\alpha^*-\alpha_k} = \frac{\|\bar{g}_k\|_2}{\|\bar{g}_{k-1}\|_2}$. Loosely speaking, it implies the increase in the risk level for next iteration from current iteration is proportional to the decrease in the norm of the gradient from previous iteration. It also ensures that the target risk level α^* is achieved when the norm of the gradient hits zero, i.e., when the algorithm converges.

In the stopping step (step 5), finding the best solution to the original CVaR optimization problem is achieved via evaluating and comparing the CVaR values at the target risk level for all the best sampled solutions found so far, and thus additional simulation budget is required; however, it is insignificant compared with the overall budget consumed.

Recall that, in GASS-CVaR-ARL, the risk level α_k is updated in accordance with the decrease in the norm of the gradient. It implies that the updating rule (4.2.1) keeps track of the algorithm's balance between the exploration and the exploitation, and makes adjustments on the risk level accordingly. Therefore, in the "warm-up" phase of the algorithm, having a small α has little negative effect on the algorithm progress since the algorithm put most of its emphasis on exploration; in the "convergence" phase of the algorithm, the risk level α_k is close to the target risk level α^*, and essentially the algorithm is solving problems close to the original one. Thus, intuitively, we expect that the number of iterations that GASS-CVaR-ARL takes to converge is similar to the one that GASS-CVaR takes to converge, which is also verified by the numerical tests presented in Chapter 4.4. Since GASS-CVaR-ARL saves simulation budget at every iteration, overall budget saving is achieved.

4.3 Convergence Analysis

Let us first analyze the convergence properties of GASS-CVaR (Algorithm 4.2.1). The analysis will rely mainly on the convergence analysis of GASS in [87] as well as the classic results in stochastic approximation methods and algorithms (see, e.g., [54], [17], [53], and [55]). The main idea is to reformulate the updating scheme on θ_k in GASS-CVaR as a generalized Robbins-Monro recursive algorithm in solving a constrained ordinary differential equation (ODE) of θ, and then show the corresponding bias term and noise term in the generalized Robbins-Monro algorithm are bounded in appropriate senses so that the sequence $\{\theta_k\}$ generated by the updating scheme converges to a limit set of the ODE w.p.1.

Following the above road map, let us first reformulate the parameter updating scheme

$$\begin{aligned}\theta_{k+1} &= \Pi_{\widetilde{\Theta}}\left\{\theta_k + \beta_k\left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)^{-1}\widehat{g}_k\right\} \\ &= \Pi_{\widetilde{\Theta}}\left\{\theta_k + \beta_k\left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)^{-1}\left(\widehat{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)]\right)\right\},\end{aligned} \tag{4.3.1}$$

in the algorithm GASS-CVaR as

$$\theta_{k+1} = \theta_k + \beta_k\left[G(\theta_k) + b_k + e_k + p_k\right]. \tag{4.3.2}$$

Here

$$\begin{aligned} G(\theta_k) &\triangleq V_k^{-1}\left(\mathbb{E}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)]\right), \\ b_k &\triangleq \widehat{V}_k^{-1}\left(\widehat{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}[\Gamma(x)]\right), \\ e_k &\triangleq \left(\widehat{V}_k^{-1} - V_k^{-1}\right)\left(\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)]\right) + V_k^{-1}\left(\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{q_k}[\Gamma(x)]\right), \end{aligned}$$

and p_k is the resulted projection error term, where for simplicity we denote

$$V_k \triangleq \left(Var_{\theta_k}[\Gamma(x)] + \epsilon I\right) \quad \text{and} \quad \widehat{V}_k \triangleq \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right).$$

In (4.3.2) the term $G(\theta_k)$ is the gradient vector field in a standard stochastic approximation algorithm, the term b_k represents the bias in estimating $\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)]$ caused

by the inexact evaluation of the shape function, the term e_k represents the Monte Carlo simulation noise in the estimators $\widehat{Var}_{\theta_k}[\Gamma(x)]$ and $\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)]$, and the term p_k represents the projection error that satisfies

$$\beta_k p_k = \theta_{k+1} - \theta_k - \beta_k \left(G(\theta_k) + b_k + e_k \right),$$

i.e., the vector that takes the current iterate back onto the constraint set $\widetilde{\Theta}$ with minimum Euclidean norm. Note that the bias term b_k is caused by both the outer-layer sampling on the solution space and the inner-layer simulation of the loss function; however, the noise term e_k only accounts for the error due to the outer-layer sampling and does not bear contribution from the inner-layer simulation.

Now let us introduce the assumptions on the algorithm and the underlying loss function for the convergence of the algorithm. The following set of assumptions is on the algorithm parameters and the choice of the exponential family of densities. It largely follows from the standard assumptions for a generalized Robbins-Monro algorithm.

Assumption 4.3.1. *(i) The step size sequence $\{\beta_k\}$ satisfies that $\beta_k > 0$ for all k, $\beta_k \searrow 0$ as $k \to \infty$, $\sum_{k=0}^{\infty} \beta_k = \infty$ and $\sum_{k=0}^{\infty} \beta_k^2 < \infty$.*

(ii) The outer-layer sample size sequence $\{N_k\}$ satisfies $N_k = N_0 \cdot k^{\tau}$ for some constant $\tau > 0$. Furthermore, the sequences $\{\beta_k\}$ and $\{N_k\}$ jointly satisfies $\frac{\beta_k}{\sqrt{N_k}} = O(k^{-\zeta})$ for some constant $\zeta > 1$.

(iii) The inner-layer sample size sequence M_k satisfies that $M_k \nearrow \infty$ as $k \to \infty$.

(iv) The sufficient statistics $\Gamma(x)$ of the chosen exponential family of densities is bounded on $\mathcal{X}$.

In the above set of assumptions, Assumption 4.3.1.(i) follows from the typical step size assumption in a gradient-based optimization algorithm. Assumption 4.3.1.(ii)

ensures that the outer-layer sample size N_k increases to infinity no slower than certain speed given a choice of the step size sequence, and it can be easily satisfied. For example, if $\beta_k = O(1/k)$, then $N_k = N_0 \cdot k^\tau$ for an arbitrary constant $\tau > 0$ is sufficient for $\frac{\beta_k}{\sqrt{N_k}} = O(k^{-\zeta})$ to hold for some constant $\zeta > 1$. Assumption 4.3.1.(iii) ensures that the error of the CVaR estimators caused by the inner-layer simulation of the loss function vanishes as $k \to \infty$. Assumption 4.3.1.(iv) is to bound those expectation and variance terms of the sufficient statistics in the algorithm. It holds for many exponential families used in practice. For example, when the solution space $\mathcal{X}$ is a nonempty compact set, the continuity of the function $\Gamma(\cdot)$ will be sufficient for Assumption 4.3.1.(iv) to hold.

The next set of assumptions is on the regularity conditions of the loss function. As noted previously, the bias term b_k is caused by the inexact evaluation of the shape function. When the shape function $S_{\theta_k}(\cdot)$ takes the form of (4.1.2), the bias is contributed by the error in estimating the $(1-\rho)$-quantile γ_{θ_k} in (4.1.3) as well as the error in the estimation of those CVaR values. Specifically, recall that for a fixed x,

$$\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x)\right) = \frac{1}{1+\exp\left(-S_0\left(-\widehat{C}_{\alpha^*}(x) - \widehat{\gamma}_{\theta_k}\right)\right)}, \tag{4.3.3}$$

where $\widehat{C}_{\alpha^*}(x)$ is the CVaR estimator given in (4.1.10), and $\widehat{\gamma}_{\theta_k}$ is the sample $(1-\rho)$-quantile of $\{-\widehat{C}_{\alpha^*}(x_k^i) : i = 1, ..., N_k\}$, i.e., $\widehat{\gamma}_{\theta_k}$ is the $(\lceil (1-\rho)N_k \rceil)^{th}$ order statistic of $\{-\widehat{C}_{\alpha^*}(x_k^i) : i = 1, ..., N_k\}$. Loosely speaking, γ_{θ_k} could be viewed as the $(1-\rho)$-level Value-at-Risk (VaR) of $-C_{\alpha^*}(x)$ w.r.t. the sampling distribution $f(x;\theta_k)$ on $\mathcal{X}$; thus, $\widehat{\gamma}_{\theta_k}$ could be regarded as a nested risk estimator where the outer-layer simulation is to estimate VaR and the inner-layer is to estimate CVaR, respectively. Therefore, bounding the bias term b_k depends on bounding the errors of both the nested risk estimator $\widehat{\gamma}_{\theta_k}$ and the one-layer CVaR estimator $\widehat{C}_{\alpha^*}(x)$. Here we will resort to the asymptotic analysis of nested risk estimators in [40] and Chapter 3.2.

To this end, let us rewrite the CVaR estimator $\widehat{C}_{\alpha^*}(x)$ in (4.1.10) as

$$\widehat{C}_{\alpha^*}(x) = C_{\alpha^*}(x) + \frac{1}{\sqrt{M_k}} \cdot \mathcal{E}_k(x), \quad \forall x \in \mathcal{X},$$

where $\mathcal{E}_k(x)$ is the standardized error (so that $\mathcal{E}_k(x)$ has a limiting distribution as $k \to \infty$ under appropriate regularity conditions). Then we impose the following set of assumptions on the loss function $l(x, \xi_x)$. It largely follows from Assumption 1 in [40] and Assumption 3.2.2 in Chapter 3.2.

Assumption 4.3.2. *(i) For all $x \in \mathcal{X}$, the loss function $l(x, \xi_x)$ has finite second moment; moreover, for all $\theta \in \widetilde{\Theta}$, the CVaR function $C_{\alpha^*}(x)$, which is a random variable under the distribution $f(\cdot; \theta)$ on $\mathcal{X}$, has finite second moment.*

(ii) For all $\theta \in \widetilde{\Theta}$ and each k, the joint density $d_k(c, e)$ of $(C_{\alpha^*}(x), \mathcal{E}_k(x))$, and its partial derivatives $\frac{\partial}{\partial c} d_k(c, e)$ and $\frac{\partial^2}{\partial c^2} d_k(c, e)$ exist for all pairs of (c, e).

(iii) For all $\theta \in \widetilde{\Theta}$ and each k, there exist nonnegative functions $D_{0,k}(\cdot)$, $D_{1,k}(\cdot)$ and $D_{2,k}(\cdot)$ such that $d_k(c, e) < D_{0,k}(e)$, $\frac{\partial}{\partial c} d_k(c, e) < D_{1,k}(e)$, and $\frac{\partial^2}{\partial c^2} d_k(c, e) < D_{2,k}(e)$ for all (c, e). Furthermore, for all $\theta \in \widetilde{\Theta}$, $\sup_k \int |e|^r D_{i,k}(e) de < \infty$ for $i = 0, 1, 2$, and $0 \leq r \leq 4$.

In the above assumption, Assumption 4.3.2.(i) ensures that a one-layer VaR/CVaR estimator defined in (4.1.10) or (4.1.11) is strongly consistent and asymptotic normally distributed, and thus the standardized estimation error $\mathcal{E}_k(x)$ has a limiting distribution as $k \to \infty$. Assumption 4.3.2.(ii) and 4.3.2.(iii) further ensure that both the p.d.f. and the c.d.f. of $\widehat{C}_{\alpha^*}(x)$ converge to the p.d.f. and the c.d.f. of $C_{\alpha^*}(x)$ sufficiently fast as $k \to \infty$, respectively. This will imply the strong consistence of the nested risk estimator $\widehat{\gamma}_{\theta_k}$ and further the convergence of the approximate shape function $\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x)\right)$, as presented in the following lemma.

Lemma 4.3.1. *Suppose the shape function $S_{\theta_k}(\cdot)$ takes the form*

$$S_{\theta_k}\left(-C_{\alpha^*}(x)\right) = \frac{1}{1 + \exp\left(-S_0\left(-C_{\alpha^*}(x) - \gamma_{\theta_k}\right)\right)},$$

where S_0 is a large positive constant, and

$$\gamma_{\theta_k} \triangleq \sup_r \left\{ r : P_{f(\cdot;\theta_k)} \left\{ x \in \mathcal{X} : -C_{\alpha^*}(x) \geq r \right\} \geq \rho \right\}$$

is the $(1-\rho)$-quantile of $(-C_{\alpha^*}(x))$ w.r.t. $f(\cdot;\theta_k)$. Further suppose that $S_{\theta_k}(-C_{\alpha^*}(x))$ is approximated by $\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x)\right)$ in (4.3.3). Then under Assumption 4.3.1.(ii), 4.3.1.(iii) and Assumption 4.3.2, we have

$$\lim_{k\to\infty} \left| \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x)\right) - S_{\theta_k}\left(-C_{\alpha^*}(x)\right) \right| = 0, \quad w.p.1, \quad \forall x \in \mathcal{X}. \tag{4.3.4}$$

The main idea of proof is to show $\widehat{C}_{\alpha^*}(x) \to C_{\alpha^*}(x)$ w.p.1 and $\widehat{\gamma}_{\theta_k} \to \gamma_{\theta_k}$ w.p.1 as $k \to \infty$. The detailed proof is included in the appendix.

Following the road map and based on Lemma 1, we next show that the bias term b_k converges to zero w.p.1. as $k \to \infty$, as presented in Lemma 2 below.

Lemma 4.3.2. *Under Assumption 4.3.1 and Assumption 4.3.2, we have*

$$\lim_{k\to\infty} \|b_k\|_2 = 0, \ w.p.1, \tag{4.3.5}$$

where recall that $\|b_k\|_2$ is the vector Euclidean norm of b_k.

The proof of Lemma 4.3.2 is included in the appendix. Continuing the road map, we next show that the summed tail error goes to zero w.p.1, as presented in the following lemma.

Lemma 4.3.3. *Under Assumption 4.3.1, we have*

$$\lim_{k\to\infty} \left\{ \sup_{n:0\leq\sum_{i=k}^{n-1}\beta_i\leq T} \left\| \sum_{i=k}^{n} \beta_i e_i \right\|_2 \right\} = 0, \quad w.p.1 \tag{4.3.6}$$

for all $T \geq 0$.

Lemma 4.3.3 is identical to Lemma 2 in [87], so we omit the proof here. With the above lemmas, we now proceed to the main result on the convergence of Algorithm 4.2.1.

Given an arbitrary $\theta \in \widetilde{\Theta}$, a set $\mathcal{C}(\theta)$ is defined as follows. For θ that lies in the interior of $\widetilde{\Theta}$, let $\mathcal{C}(\theta) = \{0\}$; for θ that lies on the boundary of $\widetilde{\Theta}$, let $\mathcal{C}(\theta)$ be the infinite convex cone generated by the outer normals at θ of the faces on which θ lies (see, e.g., [53] pp. 89). Then the updating scheme (4.3.2) in GASS-CVaR could be viewed as a noisy discretization of a constrained ODE for $\{\theta(t) : t \geq 0\}$:

$$\dot{\theta}(t) = G(\theta(t)) + p(t),\ p(t) \in -\mathcal{C}(\theta(t)),\ t \geq 0, \tag{4.3.7}$$

where $p(t)$ is the minimum force to take $\theta(t)$ back to the set $\widetilde{\Theta}$. Using the ODE approach for the convergence of the Robbins-Monro Algorithm (see, e.g., [53]), we can show that the sequence $\{\theta_k\}$ generated by (4.3.1) converges to a solution of the ODE (4.3.7). In particular, we have the following theorem.

Theorem 4.3.1. ***Convergence of GASS-CVaR****. Suppose Assumption 4.3.1 and Assumption 4.3.2 hold. Then the sequence $\{\theta_k\}$ generated by (4.3.1) converges to a limit set of the ODE (4.3.7) w.p.1. Furthermore, if the limit sets of (4.3.7) are isolated equilibrium points, then $\{\theta_k\}$ converges to a unique equilibrium point w.p.1.*

Theorem 4.3.1 is a direct consequence of Theorem 2 in [53] with Lemma 4.3.2 and Lemma 4.3.3 above. Starting from the convergence of GASS-CVaR, we will show the convergence of the algorithm GASS-CVaR-ARL. The intuition is as follows.

Recall that the updating scheme on θ in GASS-CVaR-ARL is

$$\begin{aligned}\theta_{k+1} &= \Pi_{\widetilde{\Theta}}\left\{\theta_k + \beta_k \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)^{-1} \bar{g}_k\right\} \\ &= \Pi_{\widetilde{\Theta}}\left\{\theta_k + \beta_k \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)^{-1} \left(\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{\theta_k}[\Gamma(x)]\right)\right\}.\end{aligned} \tag{4.3.8}$$

Compared with the updating scheme (4.3.1) on θ in GASS-CVaR, we could see that the approximate expectation term $\widehat{\mathbb{E}}_{q_k}[\Gamma(x)]$ in (4.3.1) is replaced by $\overline{\mathbb{E}}_{q_k}[\Gamma(x)]$ in (4.3.8) in estimating the gradient g_k. Note that the updating scheme for the risk level α_k in (4.2.1) guarantees that α_k is non-decreasing and bounded above by the target risk level α^*. Thus, the limit of the risk level sequence $\{\alpha_k\}$ exists. If we are

able to show that the limit is α^*, then the difference between $\overline{\mathbb{E}}_{q_k}[\Gamma(x)]$ and $\widehat{\mathbb{E}}_{q_k}[\Gamma(x)]$, i.e., the difference between $\bar{g}_k$ and $\widehat{g}_k$, will vanish as $k \to \infty$. The reason is that the normalized weights $\{\overline{w}_k\}$ for computing $\overline{\mathbb{E}}_{q_k}[\Gamma(x)]$ will asymptotically approach $\{\widehat{w}_k\}$ for computing $\widehat{\mathbb{E}}_{q_k}[\Gamma(x)]$ as $k \to \infty$.

Assume by contradiction that $\lim_{k\to\infty} \alpha_k = \overline{\alpha}^* < \alpha^*$. On the one hand, following from the above argument, GASS-CVaR-ARL asymptotically approaches GASS-CVaR as if the function to be minimized is $C_{\overline{\alpha}^*}(x)$ instead of $C_{\alpha^*}(x)$. Therefore, it is convergent, and thus the gradient sequence $\{g_k\}$ approaches zero w.p.1. One the other hand, the sequence $\{\|\bar{g}_k\|_2\}$ generated by (4.2.1) will always be above a certain positive value w.p.1 (otherwise α_k will converge to α^*). This contradicts with the fact that $\{g_k\}$ approaches zero w.p.1. We formalize the above analysis in the following Theorem 4.3.2. The detailed proof is included in the appendix.

Theorem 4.3.2. ***Convergence of GASS-CVaR-ARL****. Suppose Assumption 4.3.1 and Assumption 4.3.2 hold. Then the risk level sequence $\{\alpha_k\}$ generated by (4.2.1) converges to the target risk level α^* w.p.1, and the sequence $\{\theta_k\}$ generated by (4.3.8) converges to a limit set of the ODE (4.3.7) w.p.1. Furthermore, if the limit sets of (4.3.7) are isolated equilibrium points, then $\{\theta_k\}$ converges to a unique equilibrium point w.p.1.*

4.4 Numerical Experiments

We carry out numerical tests to compare the performance of GASS-CVaR and GASS-CVaR-ARL. In particular, the loss functions tested are listed in the following, among which some are designed by adding Gaussian noises to the continuous benchmark functions in [50]. However, we point out our algorithms do not have much assumption on the structure of the loss function as well as the noise. For convenience, let $\mathcal{N}(0,1)$ be a standard one-dimensional Gaussian distribution, and the loss function is in the

form of

$$l_i(x,\xi_x) = L_i(x) + \begin{cases} \sqrt{1+100\sum_{d=1}^{D}(x_d-1)^2}\cdot\mathcal{N}(0,1),\ i=0,1,3,4, \\ \sqrt{1+100\sum_{d=1}^{D}(x_d-2)^2}\cdot\mathcal{N}(0,1),\ i=2,5, \end{cases} \tag{4.4.1}$$

where D is the dimension of the solution space. Specifically, $L_0 = \sum_{d=1}^{D} x_d^2$; L_1 and L_2 are respectively Powell function and Rosenbrock function, which are badly scaled; L_3 is Rastrigin function, which is multimodal with a large number of local optima; L_4 and L_5 are respectively Pintér function and Levy function, which are badly-scaled as well as multimodal. The explicit expressions of L_i's are listed as follows, and we test all functions with $D = 10$.

(0) $L_0(x) = \sum_{d=1}^{D} x_d^2$.

(1) Powell function.

$$\begin{aligned} L_1(x) &= \sum_{d=2}^{D-2}\big[(x_{d-1}+10x_d)^2 + 5(x_{d+1}-x_{d+2})^2 + (x_d-2x_{d+1})^4 \\ &\quad +10(x_{d-1}-x_{d+2})^4\big]. \end{aligned}$$

(2) Rosenbrock function.

$$L_2(x) = \sum_{d=1}^{D-1}\left[(x_d-1)^2 + 100(x_d^2-x_{d+1})^2\right].$$

(3) Rastrigin function.

$$L_3(x) = \sum_{d=1}^{D}(x_d^2 - 10\cos(2\pi x_d)) - 10D - 1.$$

(4) Pintér function.

$$\begin{aligned} L_4(x) &= \Big[\sum_{d=1}^{D} dx_d^2 + \sum_{d=1}^{D} 20d\sin^2(x_{d-1}\sin x_d - x_d + \sin x_{d+1}) \\ &\quad + \sum_{d=1}^{D} d\log_{10}(1+d(x_{d-1}^2 - 2x_d + 3x_{d+1} - \cos x_d + 1)^2)\Big]. \end{aligned}$$

(5) Levy function.

$$\begin{aligned} L_5(x) &= -\sin^2(\pi y_1) - \sum_{d=1}^{D-1} \left[(y_d - 1)^2(1 + 10\sin^2(\pi y_d + 1))\right] \\ &\quad -(y_D - 1)^2(1 + 10\sin^2(2\pi y_D)), \end{aligned}$$

where $y_d = 1 + (x_d - 1)/4,\ d = 1, ..., D.$

Notice that if $C_{\alpha^*=0}(x) = L_i(x) = \mathbb{E}_{\xi_x}[l_i(x, \xi_x)]$ is of interest, then evidently $x^\circ = [0, ..., 0]_d$ is the minimizer for $i = 0, 1, 3, 4$, and $x^\circ = [1, ..., 1]_d$ is the minimizer for $i = 2, 5$. As the risk level of interest α^* increases, the minimizer of $C_{\alpha^*}(x)$, might be very different from x°. Specifically, the loss distribution of l_i has a relatively large variance at x° (note that it has the smallest variance at $x = [x_1^o + 1, ..., x_D^o + 1]$). This indicates that, as the risk level of interest α^* increases, the minimizer of $C_{\alpha^*}(x)$ will deviate away from x° and move towards $x = [x_1^o + 1, ..., x_D^o + 1]$, where the loss function is exposed to the lowest amount of uncertainty. Note that when $\alpha^* > 0$, the minimizers of $C_{\alpha^*}(x)$ and the minimum CVaR function values are not analytically available for the loss functions listed in (4.4.1), except for l_0.

In all the implementations, we use independent multivariate normal distribution $\mathcal{N}(\mu_k, \Sigma_k)$ as the parameterized sampling distribution $f(x; \theta_k)$ at iteration k, where $\mu_k = (\mu_k^1, ..., \mu_k^D)^T$ is the mean parameter and $\Sigma_k = diag((\sigma_k^1)^2, ..., (\sigma_k^D)^2)$ is the covariance matrix. Thus, $\theta_k = (\mu_k^1, ..., \mu_k^D; (\sigma_k^1)^2, ..., (\sigma_k^D)^2)^T$. The initial mean parameter μ_0 are drawn randomly from the uniform distribution $U[-30, 30]^D$, and the initial covariance matrix Σ_0 is set to be $\Sigma_0 = 1000 I_{D\times D}$, where $I_{D\times D}$ is the identity matrix of dimension D. From the experiment results, we notice that the performance of the algorithms is insensitive to the initial mean parameter as long as the initial covariance matrix is sufficiently large.

At iteration k, we use the shape function $S_{\theta_k}(\cdot)$ in the form of expression (4.1.2) with $S_0 = 10^5$ and $\rho = 0.1$ in (4.1.3). The $(1 - \rho)$-quantile γ_{θ_k} is estimated by the $(1 - \rho)$ sample quantile of the CVaR estimates for all the candidate solutions

generated at this iteration. The risk level of interest is $\alpha^* = 0.99$, and in GASS-CVaR-ARL the initial risk level is set to be $\alpha_0 = 0$. The sample size of candidate solutions drawn from the sampling distribution is set to be $N_k = 1000$, and the simulation budget used to estimate the CVaR of the loss function is set in a way such that the effective simulation budget is $(1 - \alpha_k)M_k = 50$. Therefore, in GASS-CVaR $M_k = 50/(1 - \alpha^*) = 50/0.01 = 5 \times 10^3$ for all k, and in GASS-CVaR-ARL $M_k = 50/(1-\alpha_k)$ at iteration k with initial simulation budget $M_0 = 50/(1-0) = 50$. The small positive constant ϵ used to ensure the positive definiteness of the Hessian is set to be $\epsilon = 10^{-10}$, and the step size β_k is set to be $\beta_k = 50/(k + 2000)^{0.6}$, which satisfies the assumptions in step 1 of both two algorithms.

We run both algorithms 50 times independently and summarize their average performance in Figure 4.4.1. Recall that, except for the loss function l_0, the minimum CVaR value is not readily available for any other loss function. So we implement GASS-CVaR with constant sample size $N = 10^3$ and simulation budget $M = 10^5$ to find them. In the upper-left plot of Figure 4.4.1 for the loss function l_0, the y-axis represents the ratio of the best CVaR values obtained by the algorithms to the true minimum CVaR value at the target risk level α^*; for all the rest of the plots, the y-axis represents the same ratio, except that the true minimum is replaced by the smallest CVaR values obtained from implementing GASS-CVaR with sample size N and simulation budget M. We observe that both algorithms (GASS-CVaR and GASS-CVaR-ARL) perform well in finding optimal solutions and minimum CVaR values. Moreover, GASS-CVaR-ARL converges faster and often reduces the total number of function evaluations needed for convergence by 2-4 times, which demonstrates the advantage of using adaptive risk level in GASS-CVaR-ARL.

Figure 4.4.2 includes two plots for the loss function l_0: the left one plots the ratio of the CVaR values evaluated at the means of the sampling distributions to the true minimum CVaR value; the right one plots the trajectory of the risk level α_k.

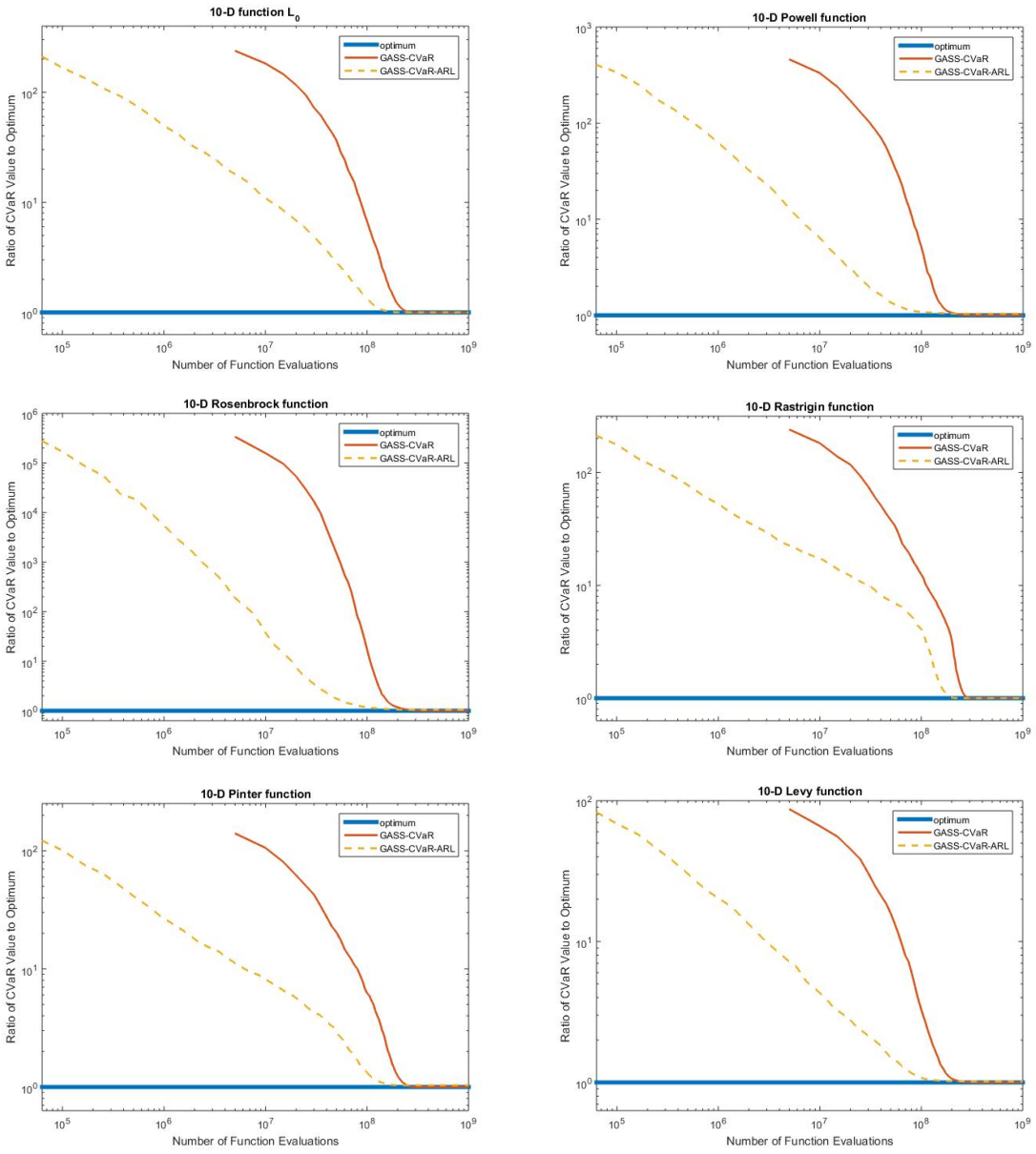

Figure 4.4.1: Average Performance of GASS-CVaR and GASS-CVaR-ARL.

We can see that the means of the sampling distributions in both GASS-CVaR and GASS-CVaR-ARL converge to the optimal solution, and GASS-CVaR-ARL achieves a faster convergence speed. Moreover, the risk level α_k in GASS-CVaR-ARL increases steadily to the target risk level $\alpha^* = 0.99$, which indicates that the norm of the gradient decreases steadily to zero and the algorithm converges.

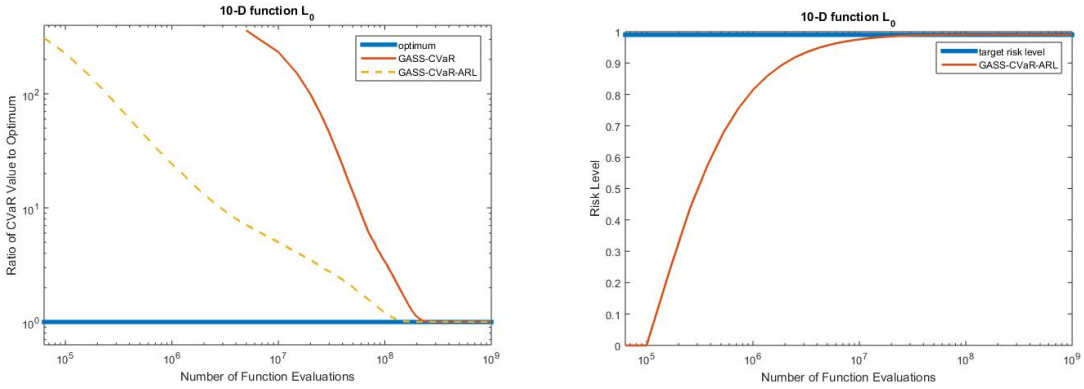

Figure 4.4.2: CVaR at Mean of the Sampling Distribution and Trajectory of Risk Level.

4.5 Conclusion and Future Research

In this chapter, we study the extension of the recently proposed GASS algorithm for deterministic non-differentiable optimization to the simulation optimization of risk measures such as VaR and CVaR. Instead of optimizing VaR/CVaR at the risk level of interest directly, we propose to initialize the algorithm at a small risk level, and then increase the risk level at each iteration adaptively such that the target risk level is achieved and the algorithm converges simultaneously. It enables us to adaptively reduce the number of samples needed to estimate VaR/CVaR at each iteration, and eventually improves the overall efficiency of the algorithm.

Directions of future research include: (1) generalizing the proposed algorithm to optimization of general coherent risk measures, where the dual representations of coherent risk measures might be useful; (2) joint sampling of the candidate solutions and loss distribution. Next let us investigate the research directions in details.

For a general coherent measure, its dual representation is equivalent to maximization of certain expectation over a constraint set of probability measures. Given a specific coherent risk measure, suppose the structure of the uncertainty set is known. Is it possible to optimize the risk measure with adaptive adjustments of the uncertainty set? As for the second direction, the relationship structure of the loss distribution

across different solutions could be explored in certain way (e.g., stochastic kriging is a classic approach to model such a relationship structure). Since multiple candidate solutions need to be drawn at each iteration, maybe the sampling of the loss distribution for the candidate solutions could be accomplished jointly using the studied relationship structure, which will significantly reduce the total number of samples required and improve the overall efficiency.

APPENDIX A

ADDITIONAL RESULTS IN CHAPTER II

The performances of the regression-based penalties under different choices of factor persistence Φ. In particular,

$$\Phi_1 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.5 \end{bmatrix}, \Phi_2 = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.7 \end{bmatrix}, \Phi_3 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.3 \end{bmatrix}, \Phi_4 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Table A.0.1: Dual Bounds of PLQC Policy with Different Φ.

Φ	D	T	Lower Bound PLQC Policy ($\$k$)	UpperBound 1 Zero Penalty ($\$k$)	UpperBound 2 First-Order ($\$k$)	UpperBound 3 Second-Order ($\$k$)	DualGap
Φ_1	1	12	7.564	8.327	7.760	7.732	2.22%
			(0.026)	(0.448)	(0.070)	(0.046)	
Φ_1	5	12	33.571	36.675	33.692	33.670	0.29%
			(0.122)	(1.890)	(0.191)	(0.103)	
Φ_1	10	12	59.391	63.451	59.893	59.837	0.75%
			(0.230)	(3.509)	(0.291)	(0.136)	
Φ_1	25	12	110.277	131.213	112.164	113.531	1.71%
			(0.506)	(8.102)	(0.552)	(0.326)	
Φ_2	1	12	7.964	9.311	8.155	8.170	2.40%
			(0.026)	(0.815)	(0.066)	(0.042)	
Φ_2	5	12	35.460	39.136	35.955	35.758	0.84%
			(0.122)	(1.787)	(0.241)	(0.096)	
Φ_2	10	12	63.287	70.992	63.770	63.885	0.76%
			(0.228)	(3.658)	(0.261)	(0.117)	
Φ_2	25	12	119.974	135.117	123.230	122.642	2.22%
			(0.505)	(9.307)	(0.526)	(0.277)	
Φ_3	1	12	2.565	3.917	2.791	2.849	8.81%
			(0.014)	(0.222)	(0.063)	(0.049)	
Φ_3	5	12	9.776	13.097	10.072	10.374	3.02%
			(0.070)	(1.217)	(0.150)	(0.084)	
Φ_3	10	12	13.946	20.531	14.388	14.489	3.17%
			(0.135)	(2.373)	(0.193)	(0.107)	
Φ_3	25	12	6.897	14.942	7.392	7.372	6.89%
			(0.327)	(5.140)	(0.250)	(0.142)	
Φ_4	1	12	1.274	2.029	1.463	1.445	13.42%
			(0.010)	(0.166)	(0.040)	(0.026)	
Φ_4	5	12	3.609	5.978	3.715	3.756	2.94%
			(0.051)	(0.817)	(0.081)	(0.043)	
Φ_4	10	12	2.638	8.173	2.741	2.707	2.62%
			(0.099)	(1.657)	(0.162)	(0.064)	
Φ_4	25	12	-18.907	-5.989	-18.356	-18.448	2.43%
			(0.245)	(3.906)	(0.159)	(0.081)	

The performances of the regression-based penalties with different choices of transaction cost parameter λ. In particular, we let

$$\lambda_1 = 1.07 \times 10^{-5},\ \lambda_2 = 2.67 \times 10^{-5},\ \lambda_3 = 3.21 \times 10^{-5},\ \lambda_4 = 4.28 \times 10^{-5}.$$

Table A.0.2: Dual Bounds of PLQC Policy with Different λ.

λ	D	T	Lower Bound PLQC Policy ($\$k$)	UpperBound 1 Zero Penalty ($\$k$)	UpperBound 2 First-Order ($\$k$)	UpperBound 3 Second-Order ($\$k$)	DualGap
λ_1	1	12	3.757	4.691	4.104	4.027	7.19%
			(0.015)	(0.250)	(0.078)	(0.056)	
λ_1	5	12	16.963	21.195	17.210	17.353	1.46%
			(0.075)	(1.324)	(0.157)	(0.094)	
λ_1	10	12	29.899	36.332	30.377	29.749	1.60%
			(0.145)	(2.444)	(0.226)	(0.104)	
λ_1	25	12	55.321	65.456	56.405	55.902	1.05%
			(0.338)	(6.056)	(0.343)	(0.201)	
λ_2	1	12	3.583	4.286	3.759	3.756	4.83%
			(0.015)	(0.234)	(0.045)	(0.034)	
λ_2	5	12	14.119	18.721	14.437	14.433	2.25%
			(0.072)	(1.221)	(0.090)	(0.040)	
λ_2	10	12	21.854	25.148	22.444	22.458	2.70%
			(0.136)	(2.096)	(0.120)	(0.062)	
λ_2	25	12	21.162	28.615	21.743	21.862	2.75%
			(0.320)	(5.446)	(0.200)	(0.117)	
λ_3	1	12	3.533	4.552	3.626	3.642	2.63%
			(0.015)	(0.251)	(0.039)	(0.023)	
λ_3	5	12	13.326	18.707	13.694	13.649	2.42%
			(0.072)	(1.083)	(0.096)	(0.039)	
λ_3	10	12	19.759	26.425	20.091	20.254	1.68%
			(0.134)	(2.290)	(0.099)	(0.045)	
λ_3	25	12	10.798	20.543	11.275	11.164	3.39%
			(0.318)	(4.439)	(0.161)	(0.085)	
λ_4	1	12	3.417	4.486	3.523	3.510	2.72%
			(0.015)	(0.273)	(0.035)	(0.018)	
λ_4	5	12	12.008	16.255	12.275	12.241	1.94%
			(0.070)	(1.120)	(0.079)	(0.031)	
λ_4	10	12	15.811	19.592	16.215	16.182	2.35%
			(0.132)	(2.070)	(0.092)	(0.043)	
λ_4	25	12	−9.375	3.890	−9.258	−9.335	0.43%
			(0.315)	(4.766)	(0.155)	(0.108)	

APPENDIX B

PROOFS OF THEOREMS IN CHAPTER III

B.1 Proof of Theorem 3.2.1

For simplicity, let us use v_α, c_α, $\widehat{v}_\alpha^N$, $\widehat{c}_\alpha^N$, $\widetilde{v}_\alpha^{N,M}$, and $\widetilde{c}_\alpha^{N,M}$ to denote VaR_α, $CVaR_\alpha$, $\widehat{VaR}_\alpha$, $\widehat{CVaR}_\alpha$, $\widetilde{VaR}_\alpha$, and $\widetilde{CVaR}_\alpha$, respectively. Therefore, we need to show that

$$\lim_{N\to\infty}\lim_{M\to\infty}\widetilde{v}_\alpha^{N,M}=v_\alpha,\quad w.p.1,\quad\text{and}\quad\lim_{N\to\infty}\lim_{M\to\infty}\widetilde{c}_\alpha^{N,M}=c_\alpha,\quad w.p.1. \tag{B.1.1}$$

In view of the error decomposition

$$\widetilde{v}_\alpha^{N,M}-v_\alpha=\left(\widetilde{v}_\alpha^{N,M}-\widehat{v}_\alpha^N\right)+\left(\widehat{v}_\alpha^N-v_\alpha\right)\quad\text{and}\quad\widetilde{c}_\alpha^{N,M}-c_\alpha=\left(\widetilde{c}_\alpha^{N,M}-\widehat{c}_\alpha^N\right)+\left(\widehat{c}_\alpha^N-c_\alpha\right), \tag{B.1.2}$$

it is sufficient to show that

$$\lim_{N\to\infty}\left(\widehat{v}_\alpha^N-v_\alpha\right)=0,\quad w.p.1.\quad\text{and}\quad\lim_{N\to\infty}\left(\widehat{c}_\alpha^N-c_\alpha\right)=0,\quad w.p.1. \tag{B.1.3}$$

and for fixed N and $\theta_1,...,\theta_N$,

$$\lim_{M\to\infty}\left(\widetilde{v}_\alpha^{N,M}-\widehat{v}_\alpha^N\right)=0,\quad w.p.1.\quad\text{and}\quad\lim_{M\to\infty}\left(\widetilde{c}_\alpha^{N,M}-\widehat{c}_\alpha^N\right)=0,\quad w.p.1. \tag{B.1.4}$$

To establish (B.1.3), we need the following lemma.

Lemma B.1.1. *Under Assumption 3.2.1.(ii),*

$$\left(\widehat{v}_\alpha^N-v_\alpha\right)=\frac{1}{f(v_\alpha)}\left(\alpha-\frac{1}{N}\sum_{i=1}^N\mathbb{1}\{H(\theta_i)\le v_\alpha\}\right)+A_N, \tag{B.1.5}$$

$$\left(\widehat{c}_\alpha^N-c_\alpha\right)=\left(\frac{1}{N}\sum_{i=1}^N\left[v_\alpha+\frac{1}{1-\alpha}\left(H(\theta_i)-v_\alpha\right)^+\right]-c_\alpha\right)+B_N, \tag{B.1.6}$$

where $A_N=O_{a.s.}(N^{-3/4}(\log N)^{3/4})$, $B_N=O_{a.s.}(N^{-1}\log N)$, *and* $(x)^+\triangleq\max\{x,0\}$. *Here note that the statement* $g(N)=O_{a.s.}(h(N))$ *means that* $g(N)\le C\cdot h(N)$ *almost surely for some constant* C.

Proof. The asymptotical representation (B.1.5) is exactly Theorem 2.5.1 in [77] under Assumption 3.2.1.(ii). The asymptotical representation (B.1.6) is the special case of Theorem 2 in [80], when the importance sampling measure $\mathcal{L} \equiv 1$. □

Notice that $\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{H(\theta_i) \leq v_\alpha\}$ is an unbiased sample estimator of α. By Strong Law of Large Numbers (SLLN),

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{H(\theta_i) \leq v_\alpha\} - \alpha = 0, \quad w.p.1. \tag{B.1.7}$$

Combining with the fact $\lim_{N\to\infty} A_N = 0,\ w.p.1,\ \lim_{N\to\infty} \left(\widehat{v}_\alpha^N - v_\alpha\right) = 0,\ w.p.1$. To show the latter half of (B.1.3), notice that $\frac{1}{N}\sum_{i=1}^{N} \left[v_\alpha + \frac{1}{1-\alpha}\left(H(\theta_i) - v_\alpha\right)^+\right]$ is an unbiased sample estimator of c_α. Furthermore, by Assumption 3.2.1.(i),

$$\mathbb{E}[H^2(\theta)] = \mathbb{E}[\mathbb{E}^2[h(\theta;\xi)|\theta]] = \int \mathbb{E}^2[h(\theta;\xi)|\theta] f(\theta) d\theta \leq \int \mathbb{E}[h^2(\theta;\xi)|\theta] f(\theta) d\theta < \infty.$$

Therefore, $Var(H(\theta))$ is finite and $Var(v_\alpha + \frac{1}{1-\alpha}\left(H(\theta) - v_\alpha\right)^+)$ is also finite. By SLLN,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \left[v_\alpha + \frac{1}{1-\alpha}\left(H(\theta_i) - v_\alpha\right)^+\right] - c_\alpha = 0, \quad w.p.1. \tag{B.1.8}$$

Combining with the fact $\lim_{N\to\infty} B_N = 0,\ w.p.1,\ \lim_{N\to\infty} \left(\widehat{c}_\alpha^N - c_\alpha\right) = 0,\ w.p.1$. To this end, (B.1.3) has been established.

It remains to establish (B.1.4) for fixed N and scenarios $\theta_1, ..., \theta_N$. That is, we need to show for fixed N and scenarios $\theta_1, ..., \theta_N$,

$$\lim_{M\to\infty} \widehat{H}_M(\theta^{(\alpha N)}) - H(\theta_{(\alpha N)}) = 0, \quad w.p.1, \tag{B.1.9}$$

$$\lim_{M\to\infty} \left(\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} \widehat{H}_M(\theta^{(i)}) - \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} H(\theta_{(i)})\right) = 0, \quad w.p.1. \tag{B.1.10}$$

Recall that for any $\theta_i, i = 1, ..., N$, $\mathbb{E}[h(\theta_i;\xi)|\theta_i] = H(\theta_i)$ and $Var[h(\theta_i;\xi)|\theta_i] = \tau_i^2 < \infty$, where we use τ_i^2 to denote $\tau_{\theta_i}^2$ with slight abuse of notations. By SLLN, we have for $i = 1, ...N$, $\widehat{H}_M(\theta_i) \overset{M\to\infty}{\to} H(\theta_i),\ w.p.1$. Let $\Omega_i \subseteq \Omega$ be the set of such convergent

scenarios for $i = 1, ..., N$, where Ω is the underlying sample space. Thus $P(\Omega_i) = 1$. Denote $\bar{\Omega} := \bigcap_{i=1}^{N} \Omega_i$, the intersection of all convergent scenario sets. Clearly, by Boole's Inequality $P(\bar{\Omega}) = 1$. Let us also denote, for any scenario $w \in \bar{\Omega}$, $\widehat{H}_M^w(\theta)$ as the sample realization of $\widehat{H}_M(\theta)$, $i = 1, .., N$. Therefore, $\forall w \in \bar{\Omega}$

$$\lim_{M\to\infty} (\widehat{H}_M^w(\theta_1), ..., \widehat{H}_M^w(\theta_N)) = (H(\theta_1), ..., H(\theta_N)). \tag{B.1.11}$$

Let $\epsilon := \frac{1}{3} \min\{H(\theta_i) - H(\theta_j) : \ i \neq j, \ i, j = 1, ..., N\}$. By definition, (B.1.11) implies that there exists a sufficient large M_ϵ such that $\forall M \geq M_\epsilon$, $|\widehat{H}_M^w(\theta_i) - H(\theta_i)| < \epsilon, i = 1, ..., N$. It follows that, $\forall M \geq M_\epsilon$,

$$\widehat{H}_M^w(\theta_{(1)}) < \widehat{H}_M^w(\theta_{(2)}) < \cdots < \widehat{H}_M^w(\theta_{(N)}).$$

That is, $\forall M \geq M_\epsilon$, the sampling error so small that the order sequence of the mean response is not perturbed. Thus, $\forall M \geq M_\epsilon$, $(\theta_w^{(1)}, .., \theta_w^{(N)}) = (\theta_{(1)}, ..., \theta_{(N)})$, where $\theta_w^{(i)}$ is the sample realization of $\theta^{(i)}$ with scenario w. Therefore, for any scenario $w \in \bar{\Omega}$,

$$\lim_{M\to\infty} \widehat{H}_M^w(\theta_w^{(\alpha N)}) = \lim_{M\to\infty} \widehat{H}_M^w(\theta_{(\alpha N)}) = H(\theta_{(\alpha N)}),$$

and

$$\lim_{M\to\infty} \frac{1}{(1-\alpha)N} \sum_{i=\alpha N}^{N} \widehat{H}_M^w(\theta_w^{(i)}) = \lim_{M\to\infty} \frac{1}{(1-\alpha)N} \sum_{i=\alpha N}^{N} \widehat{H}_M^w(\theta_{(i)}) = \frac{1}{(1-\alpha)N} \sum_{i=\alpha N}^{N} H(\theta_{(i)}). \tag{B.1.12}$$

Notice $P(\bar{\Omega}) = 1$, (B.1.9) and (B.1.10) naturally hold.

B.2 Proof of Theorem 3.2.2

Recall we need to show that

$$\lim_{N,M\to\infty} \widetilde{v}_\alpha^{N,M} = v_\alpha, \quad w.p.1, \quad \text{and} \quad \lim_{N,M\to\infty} \widetilde{c}_\alpha^{N,M} = c_\alpha, \quad w.p.1. \tag{B.2.1}$$

In addition to the notations previously introduced in Appendix B.1, let us further use $\breve{v}_\alpha^M$ and $\breve{c}_\alpha^M$ to denote $VaR_\alpha(\widehat{H}_M(\theta))$ and $CVaR_\alpha(\widehat{H}_M(\theta))$, respectively. That is, $\breve{v}_\alpha^M$

and $\breve{c}_\alpha^M$ are the exact α-level VaR and CVaR of the "noised" mean response $\widehat{H}_M(\theta)$, respectively.

As mentioned after Theorem 3.2.2, in view of the fact that $\widetilde{VaR}_\alpha(H(\theta)) = \widehat{VaR}_\alpha(\widehat{H}_M(\theta))$ and $\widetilde{CVaR}_\alpha(H(\theta)) = \widehat{CVaR}_\alpha(\widehat{H}_M(\theta))$, $\widetilde{v}_\alpha^{N,M}$ and $\widetilde{c}_\alpha^{N,M}$ could be regarded as the one layer Monte Carlo estimator of $\breve{v}_\alpha^M$ and $\breve{c}_\alpha^M$, respectively. This observation inspires us to consider the following error decomposition

$$\widetilde{v}_\alpha^{N,M} - v_\alpha = \left(\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M\right) + \left(\breve{v}_\alpha^M - v_\alpha\right) \quad \text{and} \quad \widetilde{c}_\alpha^{N,M} - c_\alpha = \left(\widetilde{c}_\alpha^{N,M} - \breve{c}_\alpha^M\right) + \left(\breve{c}_\alpha^M - c_\alpha\right). \tag{B.2.2}$$

Therefore, it is sufficient to show that

$$\lim_{M\to\infty} \breve{v}_\alpha^M = v_\alpha \quad \text{and} \quad \lim_{M\to\infty} \breve{c}_\alpha^M = c_\alpha, \tag{B.2.3}$$

and **uniformly** for all M,

$$\lim_{N\to\infty} \widetilde{v}_\alpha^{N,M} = \breve{v}_\alpha^M \quad w.p.1 \quad \text{and} \quad \lim_{N\to\infty} \widetilde{c}_\alpha^{N,M} = \breve{c}_\alpha^M \quad w.p.1. \tag{B.2.4}$$

Let us first establish (B.2.3). The following lemmas will be useful.

Lemma B.2.1. *Under Assumption 3.2.2, if a sequence $t_M \to t$ as $M \to \infty$, then $\widetilde{f}_M(t_M) \to f(t)$ and $\widetilde{f}'_M(t_M) \to f'(t)$ as $M \to \infty$, where recall $\widetilde{f}_M(\cdot)$ is the p.d.f. of the "noised" mean response $\widehat{H}_M(\theta)$.*

Proof. This result is exactly Lemma 1 in [40]. For convenience, we will briefly present the proof. Recall that $\widehat{H}_M(\theta) = H(\theta) + \bar{\mathcal{E}}_M/\sqrt{M}$, where $(H(\theta), \bar{\mathcal{E}}_M)$ has a joint distribution $p_M(h, e)$. Therefore,

$$\widetilde{f}_M(t_M) = \int_{\mathbb{R}} p_M(t_M - e/\sqrt{M}, e)de \quad \text{and} \quad f(t) = \int_{\mathbb{R}} p_M(t, e)de.$$

It follows that

$$\widetilde{f}_M(t_M) - f(t) = \int_{\mathbb{R}} \Big(p_M(t - e/\sqrt{M}, e) - p_M(t, e)\Big)\, de.$$

By Taylor series expansion, this equals

$$(t_M - t)\int_{\mathbb{R}} \frac{\partial}{\partial t} p_M(\check{t}_M, e)de - \frac{1}{\sqrt{M}}\int_{\mathbb{R}} e\frac{\partial}{\partial t} p_M(\check{t}_M, e)de,$$

where $\check{t}_M$ lives in between t_M and t. By Assumption 3.2.2 and the fact that $t_M \to t$ as $M \to \infty$, both terms converge to zero as $M \to \infty$. □

Lemma B.2.2. *Under Assumption 3.2.2,*

$$\check{v}_\alpha^M = v_\alpha + \frac{-\Lambda'(v_\alpha)}{Mf(v_\alpha)} + o_M(\frac{1}{M}), \tag{B.2.5}$$

where the function $\Lambda(t) = 1/2f(t)\mathbb{E}[\tau_\theta^2|H(\theta) = t]$ *and* $o_M(\frac{1}{M})$ *means this quantity goes to zero faster than* $\frac{1}{M}$ *(almost surely).*

Proof. This result is very similar to Proposition 1 in [40]. The proof here will mainly follow [40]'s proof.

Recall that $\widetilde{F}_M(\cdot)$ is the c.d.f. of the "noised" mean response $\widehat{H}_M(\theta)$, and $\check{v}_\alpha^M$ is the exact α-level VaR of $\widehat{H}_M(\theta)$. Thus, $\widetilde{F}_M(\check{v}_\alpha^M) = \alpha$. By Taylor expansion, we have

$$\alpha = \widetilde{F}_M(\check{v}_\alpha^M) = \widetilde{F}_M(v_\alpha) + (\check{v}_\alpha^M - v_\alpha)\widetilde{f}_M(v_\alpha) + \frac{(\check{v}_\alpha^M - v_\alpha)^2}{2}\widetilde{f}'_M(\check{v}_\alpha^M),$$

where $\check{v}_\alpha^M$ lives in between $\check{v}_\alpha^M$ and v_α. Therefore,

$$\alpha - \widetilde{F}_M(v_\alpha) = (\check{v}_\alpha^M - v_\alpha)\widetilde{f}_M(v_\alpha) + \frac{(\check{v}_\alpha^M - v_\alpha)^2}{2}\widetilde{f}'_M(\check{v}_\alpha^M), \tag{B.2.6}$$

Furthermore, notice that

$$\widetilde{F}_M(v_\alpha) = \int_{-\infty}^{v_\alpha} \widetilde{f}_M(t)dt = \int_{\mathbb{R}}\int_{-\infty}^{v_\alpha - e/\sqrt{M}} p_M(t,e)dtde, \tag{B.2.7}$$

and

$$\alpha = F(v_\alpha) = \int_{-\infty}^{v_\alpha} f(t)dt = \int_{\mathbb{R}}\int_{-\infty}^{v_\alpha} p_M(t,e)dtde. \tag{B.2.8}$$

Combining (B.2.7) and (B.2.8), we have

$$\alpha - \widetilde{F}_M(v_\alpha) = \int_{\mathbb{R}}\int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} p_M(t,e)dtde. \tag{B.2.9}$$

By Taylor expansion, we have

$$p_M(t,e) = p_M(v_\alpha,e) + (t - v_\alpha)\frac{\partial}{\partial t}p_M(v_\alpha,e) + \frac{(t-v_\alpha)^2}{2}\frac{\partial^2}{\partial t^2}p_M(\check{v}_\alpha,e), \tag{B.2.10}$$

where $\check{v}_\alpha$ lives in between v_α and t. Hence,

$$\alpha - \widetilde{F}_M(v_\alpha) = \int_{\mathbb{R}} \int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} p_M(v_\alpha, e) dt de + \int_{\mathbb{R}} \int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} (t - v_\alpha) \frac{\partial}{\partial t} p_M(v_\alpha, e) dt de + \int_{\mathbb{R}} \int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} \frac{(t - v_\alpha)^2}{2} \frac{\partial^2}{\partial t^2} p_M(\check{v}_\alpha, e) dt de. \tag{B.2.11}$$

The first term of the right hand side of (B.2.11) satisfies

$$\int_{\mathbb{R}} \int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} p_M(v_\alpha, e) dt de = \int_{\mathbb{R}} \frac{e}{\sqrt{M}} p_M(v_\alpha, e) de = \frac{f(v_\alpha)}{\sqrt{M}} \mathbb{E}[\bar{\mathcal{E}}_M | H(\theta) = v_\alpha] = 0.$$

The second term of (B.2.11) satisfies

$$\begin{aligned} & \int_{\mathbb{R}} \int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} (t - v_\alpha) \frac{\partial}{\partial t} p_M(v_\alpha, e) dt de \\ = & -\frac{1}{2M} \int_{\mathbb{R}} e^2 \frac{\partial}{\partial t} p_M(v_\alpha, e) de = -\frac{1}{2M} \frac{\partial}{\partial t} \int_{\mathbb{R}} e^2 p_M(v_\alpha, e) de \\ = & -\frac{1}{2M} \frac{\partial}{\partial t} f(v_\alpha) \mathbb{E}[\tau_\theta^2 | H(\theta) = v_\alpha] = -\frac{1}{M} \Lambda'(v_\alpha). \end{aligned}$$

By Assumption 3.2.2, the third term is in the order of $O_M(M^{-\frac{3}{2}})$. Therefore,

$$\alpha - \widetilde{F}_M(v_\alpha) = -\frac{1}{M} \Lambda'(v_\alpha) + O_M(M^{-\frac{3}{2}}). \tag{B.2.12}$$

Combining (B.2.12) with (B.2.6), we have

$$(\check{v}_\alpha^M - v_\alpha) \widetilde{f}_M(v_\alpha) + \frac{(\check{v}_\alpha^M - v_\alpha)^2}{2} \widetilde{f}'_M(\check{v}_\alpha^M) = -\frac{1}{M} \Lambda'(v_\alpha) + O_M(M^{-\frac{3}{2}}).$$

Note that by Assumption 3.2.2, it is easy to see that $\widetilde{f}'_M(t)$ is uniformly bounded for all t and M. Combining with Lemma B.2.1, (B.2.5) holds. □

Lemma B.2.3. *Under Assumption 3.2.2,*

$$\check{c}_\alpha^M = c_\alpha + \frac{\Lambda(v_\alpha)}{(1-\alpha)M} + o_M(\frac{1}{M}). \tag{B.2.13}$$

Proof. The result here is very similar to Proposition 3 in [40], and our proof will mainly follow [40]'s proof. Note that by Mean Value Theorem,

$$\check{c}_\alpha^M = \frac{1}{1-\alpha} \mathbb{E}\left[\widetilde{H}_M(\theta) \cdot \mathbb{1}\{\widetilde{H}_M(\theta) \geq \check{v}_\alpha^M\}\right] = \frac{1}{1-\alpha} \int_{\check{v}_\alpha^M}^{\infty} t \widetilde{f}_M(t) dt$$

$$
\begin{aligned}
&= \frac{1}{1-\alpha}\int_{v_\alpha}^{\infty} t\widetilde{f}_M(t)dt + \frac{1}{1-\alpha}\int_{\breve{v}_\alpha^M}^{v_\alpha} t\widetilde{f}_M(t)dt \\
&= \frac{1}{1-\alpha}\mathbb{E}\left[\widetilde{H}_M(\theta)\cdot \mathbb{1}\{\widetilde{H}_M(\theta)\geq v_\alpha\}\right] + \frac{1}{1-\alpha}(v_\alpha - \breve{v}_\alpha^M)t_v\widetilde{f}_M(t_v),
\end{aligned}
$$

where t_v lives in between v_α and $\breve{v}_\alpha^M$. By Lemma B.2.2, we know

$$
\frac{1}{1-\alpha}(v_\alpha - \breve{v}_\alpha^M)t_v\widetilde{f}_M(t_v) = \frac{v_\alpha\Lambda'(v_\alpha)}{(1-\alpha)M} + o_M(\frac{1}{M}).
$$

Therefore,

$$
\breve{c}_\alpha^M = \frac{1}{1-\alpha}\mathbb{E}\left[\widetilde{H}_M(\theta)\cdot \mathbb{1}\{\widetilde{H}_M(\theta)\geq v_\alpha\}\right] + \frac{v_\alpha\Lambda'(v_\alpha)}{(1-\alpha)M} + o_M(\frac{1}{M}). \tag{B.2.14}
$$

Further notice that

$$
\begin{aligned}
&\frac{1}{1-\alpha}\mathbb{E}\left[\widetilde{H}_M(\theta)\cdot \mathbb{1}\{\widetilde{H}_M(\theta)\geq v_\alpha\}\right] \\
&= \frac{1}{1-\alpha}\int_{\mathbb{R}}\int_{v_\alpha - e/\sqrt{M}}^{\infty}(t+e/\sqrt{M})p_M(t,e)dtde,
\end{aligned} \tag{B.2.15}
$$

$$
c_\alpha = \frac{1}{1-\alpha}\mathbb{E}\left[H(\theta)\cdot \mathbb{1}\{H(\theta)\geq v_\alpha\}\right] = \frac{1}{1-\alpha}\int_{\mathbb{R}}\int_{v_\alpha}^{\infty} tp_M(t,e)dtde, \tag{B.2.16}
$$

and

$$
\int_{\mathbb{R}}\int_{v_\alpha}^{\infty} ep_M(t,e)dtde = \int_{v_\alpha}^{\infty}\mathbb{E}[\tilde{\mathcal{E}}_M|H(\theta)=t]f(t)dt = 0.
$$

Therefore,

$$
\begin{aligned}
\breve{c}_\alpha^M - c_\alpha &= \frac{1}{1-\alpha}\left(\int_{\mathbb{R}}\int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} tp_M(t,e)dtde + \frac{1}{\sqrt{N}}\int_{\mathbb{R}} e\int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} p_M(t,e)dtde\right) \\
&\quad + \frac{v_\alpha\Lambda'(v_\alpha)}{(1-\alpha)M} + o_M(\frac{1}{M}).
\end{aligned} \tag{B.2.17}
$$

Similar to the derivation (by taking Taylor expansion) from (B.2.9) to (B.2.12), we have

$$
\frac{1}{1-\alpha}\int_{\mathbb{R}}\int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} tp_M(t,e)dtde = -\frac{\Lambda(v_\alpha)}{(1-\alpha)M} - \frac{v_\alpha\Lambda'(v_\alpha)}{(1-\alpha)M} + O_M(M^{-\frac{3}{2}}), \tag{B.2.18}
$$

and

$$
\frac{1}{1-\alpha}\frac{1}{\sqrt{N}}\int_{\mathbb{R}} e\int_{v_\alpha - e/\sqrt{M}}^{v_\alpha} p_M(t,e)dtde = 2\frac{\Lambda(v_\alpha)}{(1-\alpha)M} + O_M(M^{-\frac{3}{2}}). \tag{B.2.19}
$$

Combining (B.2.17), (B.2.18), and (B.2.19), (B.2.13) holds and Lemma B.2.3 is proven. □

By Lemma B.2.2 and Lemma B.2.3, (B.2.3) naturally holds. It remains to show (B.2.4). For simplicity, let us use $G(\cdot)$ and $\widetilde{G}_M(\cdot)$ to denote the inverse functions of $F(\cdot)$ and $\widetilde{F}_M(\cdot)$, respectively. Furthermore, denote $U(\theta) = \widetilde{F}_M(\widehat{H}_M(\theta))$. Clearly, $\widehat{H}_M(\theta) = \widetilde{G}_M(U(\theta))$ and $\breve{v}_\alpha^M = \widetilde{G}_M(\alpha)$. It is easy to see that $U(\theta)$ is uniformly distributed over $[0,1]$. Moreover, from the relationship $\widehat{H}_M(\theta_{(1)}) < \cdots < \widehat{H}_M(\theta_{(N)})$, we know that $U(\theta_{(1)}) < \cdots U(\theta_{(N)})$ is the corresponding order statistics of N i.i.d. uniformly distributed random variables. Furthermore, let us use $\widehat{F}_u^N(\cdot)$ to denote the sample c.d.f. induced by $U(\theta_1), ..., U(\theta_N)$. That is

$$\widehat{F}_u^N(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{U(\theta_i) \le t\}.$$

The following lemma is sufficient for proving (B.2.4).

Lemma B.2.4. *Under Assumption 3.2.2,*

$$(\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M) = \frac{1}{\widetilde{f}(\breve{v}_\alpha^M)}\left(\alpha - \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{\widehat{H}_M(\theta_i) \le \breve{v}_\alpha^M\}\right) + O_{a.s.}(N^{-3/4}(\log N)^{3/4}), \quad \text{(B.2.20)}$$

and

$$(\widetilde{c}_\alpha^{N,M} - \breve{c}_\alpha^M) = \left(\frac{1}{N}\sum_{i=1}^{N}\left[\breve{v}_\alpha^M + \frac{1}{1-\alpha}\left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right] - \breve{c}_\alpha^M\right) + O_{a.s.}(N^{-1}\log N), \quad \text{(B.2.21)}$$

where $O_{a.s.}(N^{-3/4}(\log N)^{3/4})$ and $O_{a.s.}(N^{-1}\log N)$ hold uniformly for all M.

Proof. Let us first establish (B.2.20). By Lemma B.1.1, we know that

$$U(\theta_{(\alpha N)}) - \alpha = \left(\alpha - \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{U(\theta_i) \le \alpha\}\right) + O_{a.s.}(N^{-3/4}(\log N)^{3/4}). \quad \text{(B.2.22)}$$

Furthermore, by Taylor expansion,

$$\begin{aligned}
\widetilde{v}_\alpha^{N,M} &= \widehat{H}_M(\theta_{(\alpha N)}) = \widetilde{G}_M(U(\theta_{(\alpha N)})) \\
&= \widetilde{G}_M(\alpha) + (U(\theta_{(\alpha N)}) - \alpha)\widetilde{G}'_M(\alpha) + \frac{(U(\theta_{(\alpha N)}) - \alpha)^2}{2}\widetilde{G}''_M(u) \\
&= \breve{v}_\alpha^M + \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}(U(\theta_{(\alpha N)}) - \alpha) + \left(-\frac{\widetilde{f}'_M(\widetilde{G}_M(u))}{\widetilde{f}_M^3(\widetilde{G}_M(u))}\right)\frac{(U(\theta_{(\alpha N)}) - \alpha)^2}{2},
\end{aligned}$$

where u lives in between $U(\theta_{(\alpha N)})$ and α, and we use the facts $\widetilde{G}_M(\alpha) = \breve{v}_\alpha^M$, $\widetilde{G}'_M(\alpha) = 1/\widetilde{f}_M(\breve{v}_\alpha^M)$, and $\widetilde{G}''_M(u) = \widetilde{f}'_M(\widetilde{G}_M(u))/\widetilde{f}^3_M(\widetilde{G}_M(u))$. Therefore,

$$\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M = \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}(U(\theta_{(\alpha N)}) - \alpha) + \left(-\frac{\widetilde{f}'_M(\widetilde{G}_M(u))}{\widetilde{f}^3_M(\widetilde{G}_M(u))}\right)\frac{(U(\theta_{(\alpha N)}) - \alpha)^2}{2}. \quad \text{(B.2.23)}$$

On the other hand, by Lemma 2.5.4B in [77], we have for sufficiently large N

$$|U(\theta_{(\alpha N)}) - \alpha| \leq 2N^{-\frac{1}{2}}(\log N)^{\frac{1}{2}}. \quad \text{(B.2.24)}$$

Combining with (B.2.22) and (B.2.23), we have

$$\begin{aligned}
&\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M \\
= \;& \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}\left\{\left(\frac{1}{N}\sum_{i=1}^N \mathbb{1}\{U(\theta_i) \leq \alpha\} - \alpha\right) + O_{a.s.}(N^{-3/4}(\log N)^{3/4})\right\} \quad \text{(B.2.25)} \\
= \;& \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}\left(\frac{1}{N}\sum_{i=1}^N \mathbb{1}\{U(\theta_i) \leq \alpha\} - \alpha\right) + \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}O_{a.s.}(N^{-3/4}(\log N)^{3/4}) \\
= \;& \frac{1}{\widetilde{f}(\breve{v}_\alpha^M)}\left(\frac{1}{N}\sum_{i=1}^N \mathbb{1}\{\widehat{H}_M(\theta_i) \leq \breve{v}_\alpha^M\} - \alpha\right) + \frac{1}{\widetilde{f}_M(\breve{v}_\alpha^M)}O_{a.s.}(N^{-3/4}(\log N)^{3/4}).
\end{aligned}$$

Notice that $\widetilde{f}_M(\breve{v}_\alpha^M)$ is strictly positive and $\widetilde{f}_M(\breve{v}_\alpha^M) \to f(v_\alpha)$ as $M \to \infty$. Therefore, $\sup_M 1/\widetilde{f}_M(\breve{v}_\alpha^M) < \infty$. Thus, (B.2.20) holds.

It remains to show (B.2.21). Notice that by definition

$$\begin{aligned}
&\widetilde{c}_\alpha^{N,M} - \breve{c}_\alpha^M = \frac{1}{(1-\alpha)N}\sum_{i=1}^N \widehat{H}_M(\theta_i)\mathbb{1}\{\widehat{H}_M(\theta_i) \geq \widetilde{v}_\alpha^{N,M}\} - \breve{c}_\alpha^M \\
= \;& \widetilde{v}_\alpha^{N,M} + \frac{1}{(1-\alpha)N}\sum_{i=1}^N \left(\widehat{H}_M(\theta_i) - \widetilde{v}_\alpha^{N,M}\right)^+ - \breve{c}_\alpha^M \\
= \;& \left(\frac{1}{N}\sum_{i=1}^N \left[\breve{v}_\alpha^M + \frac{1}{1-\alpha}\left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right] - \breve{c}_\alpha^M\right) \\
& + \left(\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M\right) + \frac{1}{(1-\alpha)N}\sum_{i=1}^N \left[\left(\widehat{H}_M(\theta_i) - \widetilde{v}_\alpha^{N,M}\right)^+ - \left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right] \\
= \;& \left(\frac{1}{N}\sum_{i=1}^N \left[\breve{v}_\alpha^M + \frac{1}{1-\alpha}\left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right] - \breve{c}_\alpha^M\right) + (*),
\end{aligned}$$

where

$$(*) := \left(\widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M\right) + \frac{1}{(1-\alpha)N}\sum_{i=1}^N \left[\left(\widehat{H}_M(\theta_i) - \widetilde{v}_\alpha^{N,M}\right)^+ - \left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right].$$

We only need to show that $(*)$ is in the order of $O_{a.s.}(N^{-1}\log N)$ uniformly for all M. Note that the second term in $(*)$

$$
\begin{aligned}
& \frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left[\left(\widehat{H}_M(\theta_i)-\widetilde{v}_\alpha^{N,M}\right)^+-\left(\widehat{H}_M(\theta_i)-\breve{v}_\alpha^{M}\right)^+\right] \\
= & \frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\Bigg[\left(\widehat{H}_M(\theta_i)-\widetilde{v}_\alpha^{N,M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\geq\widetilde{v}_\alpha^{N,M}\} \\
& -\left(\widehat{H}_M(\theta_i)-\breve{v}_\alpha^{M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\geq\breve{v}_\alpha^{M}\}\Bigg] \\
= & \frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left[\left(\breve{v}_\alpha^{M}-\widetilde{v}_\alpha^{N,M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\geq\widetilde{v}_\alpha^{N,M}\}\right] \\
& +\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left(\widehat{H}_M(\theta_i)-\breve{v}_\alpha^{M}\right)\left[\mathbb{1}\{\widehat{H}_M(\theta_i)\geq\widetilde{v}_\alpha^{N,M}\}-\mathbb{1}\{\widehat{H}_M(\theta_i)\geq\breve{v}_\alpha^{M}\}\right] \\
= & \frac{1}{(1-\alpha)N}\left(\breve{v}_\alpha^{M}-\widetilde{v}_\alpha^{N,M}\right)+\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left[\left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}\right] \\
& +\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left(\widehat{H}_M(\theta_i)-\breve{v}_\alpha^{M}\right)\left[\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\breve{v}_\alpha^{M}\}-\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}\right] \\
= & \frac{1}{(1-\alpha)N}\left(\breve{v}_\alpha^{M}-\widetilde{v}_\alpha^{N,M}\right)+\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left[\left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}\right] \\
& +(***),
\end{aligned}
$$

where

$$
(***)=\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left(\widehat{H}_M(\theta_i)-\breve{v}_\alpha^{M}\right)\left[\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\breve{v}_\alpha^{M}\}-\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}\right].
$$

Further note that

$$
\begin{aligned}
& \left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)+\frac{1}{(1-\alpha)N}\left(\breve{v}_\alpha^{M}-\widetilde{v}_\alpha^{N,M}\right) \\
& +\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\left[\left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}\right] \\
= & \frac{1}{(1-\alpha)}\left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{\widehat{H}_M(\theta_i)\leq\widetilde{v}_\alpha^{N,M}\}-\alpha\right] \\
= & \frac{1}{(1-\alpha)}\left(\widetilde{v}_\alpha^{N,M}-\breve{v}_\alpha^{M}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{U(\theta_i)\leq U(\theta_{(\alpha N)})\}-\alpha\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{(1-\alpha)} \left(\widetilde{v}_{\alpha}^{N,M} - \breve{v}_{\alpha}^{M}\right) \left(\widehat{F}_{u}^{N}(\theta_{(\alpha N)}) - \alpha\right) \\
&\stackrel{\triangle}{=} (**).
\end{aligned}
$$

Note that $(*) = (**) + (***)$, we only need to show that $(**)$ and $(***)$ both are in the order of $O_{a.s.}(N^{-1}\log N)$ uniformly for all M.

By Lemma 2.5.4B in [77], we know that for sufficiently large N (can be verified this is uniform for all M, as in (B.2.25))

$$
|\widetilde{v}_{\alpha}^{N,M} - \breve{v}_{\alpha}^{M}| \leq \frac{2}{\widetilde{f}_{M}(\breve{v}_{\alpha}^{M})} N^{-\frac{1}{2}} (\log N)^{\frac{1}{2}} . \tag{B.2.26}
$$

Moreover, by applying Theorem 2.5.1 and Lemma 2.5.4B in [77] on $U(\theta)$, we have for sufficiently large N

$$
|\widehat{F}_{u}^{N}(\alpha) - \alpha| = 2N^{-\frac{1}{2}} (\log N)^{\frac{1}{2}} + O_{a.s.}(N^{-3/4}(\log N)^{3/4}). \tag{B.2.27}
$$

Applying Lemma 2.5.4B and Lemma 2.5.4E (with $c_0 = 2$, $q = 1/2$) in [77] on $U(\theta)$, we have for sufficiently large N

$$
|\widehat{F}_{u}^{N}(\theta_{(\alpha N)}) - \widehat{F}_{u}^{N}(\alpha)| = 2N^{-\frac{1}{2}} (\log N)^{\frac{1}{2}} + O_{a.s.}(N^{-3/4}(\log N)^{3/4}). \tag{B.2.28}
$$

Combining (B.2.27) and (B.2.28), we have for sufficiently large N

$$
|\widehat{F}_{u}^{N}(\theta_{(\alpha N)}) - \alpha| \leq 4N^{-\frac{1}{2}} (\log N)^{\frac{1}{2}} + O_{a.s.}(N^{-3/4}(\log N)^{3/4}).
$$

Combining with (B.2.26), we have for sufficiently large N (uniform for all M)

$$
(**) = \frac{8}{\widetilde{f}_{M}(\breve{v}_{\alpha}^{M})} \left(N^{-1} (\log N) + O_{a.s.}(N^{-5/4}(\log N)^{5/4})\right) \tag{B.2.29}
$$

In view of the fact $\sup_{M} 1/\widetilde{f}_{M}(\breve{v}_{\alpha}^{M}) < \infty$, we have $(**)$ in the order of $O_{a.s.}(N^{-1}\log N)$ uniformly for all M. What is left is show $(***)$ is also in the order of $O_{a.s.}(N^{-1}\log N)$

uniformly for all M.

$$
\begin{aligned}
|(***)| &= \left| \frac{1}{(1-\alpha)N} \sum_{i=1}^{N} \left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M \right) \left[\mathbb{1}\{\widehat{H}_M(\theta_i) \le \breve{v}_\alpha^M\} - \mathbb{1}\{\widehat{H}_M(\theta_i) \le \widetilde{v}_\alpha^{N,M}\} \right] \right| \\
\le & \frac{1}{(1-\alpha)} \left| \widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M \right| \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{\widehat{H}_M(\theta_i) \le \breve{v}_\alpha^M\} - \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{\widehat{H}_M(\theta_i) \le \widetilde{v}_\alpha^{N,M}\} \right| \\
= & \frac{1}{(1-\alpha)} \left| \widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M \right| \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{U(\theta_i) \le \alpha\} - \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{U(\theta_i) \le U(\theta_{(\alpha N)})\} \right| \\
= & \frac{1}{(1-\alpha)} \left| \widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M \right| \left| \widehat{F}_u^N(\theta_{(\alpha N)}) - \widehat{F}_u^N(\alpha) \right|.
\end{aligned}
$$

By (B.2.26) and (B.2.28), we have for sufficiently large N (uniform for all M)

$$
\begin{aligned}
(**) &\le \frac{1}{(1-\alpha)} \left| \widetilde{v}_\alpha^{N,M} - \breve{v}_\alpha^M \right| \left| \widehat{F}_u^N(\theta_{(\alpha N)}) - \widehat{F}_u^N(\alpha) \right| \\
&= \frac{4}{\widetilde{f}_M(\breve{v}_\alpha^M)} \left(N^{-1} (\log N) + O_{a.s.}(N^{-5/4} (\log N)^{5/4}) \right). \qquad \text{(B.2.30)}
\end{aligned}
$$

Again, in view of the fact that $\sup_M 1/\widetilde{f}_M(\breve{v}_\alpha^M) < \infty$, we have $(***)$ in the order of $O_{a.s.}(N^{-1} \log N)$ uniformly for all M. Proof is complete. □

B.3 Proof of Theorem 3.2.3

Following the notations introduced in Appendix B.1 and B.2, we need to show

$$
\lim_{N\to\infty} \sqrt{N} \left(\widehat{v}_\alpha^N - v_\alpha \right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{var} \mathcal{N}(0,1) \quad \text{and} \quad \lim_{N\to\infty} \sqrt{N} \left(\widehat{c}_\alpha^N - c_\alpha \right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{cvar} \mathcal{N}(0,1), \tag{B.3.1}
$$

where

$$
\sigma_{var} \overset{\triangle}{=} \sqrt{\alpha(1-\alpha)}/f(v_\alpha) \quad \text{and} \quad \sigma_{cvar} \overset{\triangle}{=} \sqrt{Var\left[(H(\theta) - v_\alpha)^+ \right]}/(1-\alpha).
$$

Furthermore,

$$
\begin{cases}
\lim_{N\to\infty} \lim_{M\to\infty} \sqrt{M} \left(\widetilde{v}_\alpha^{N,M} - \widehat{v}_\alpha^N \right) \overset{\mathcal{D}}{\Rightarrow} \tau_{var} \mathcal{N}(0,1), \\
\lim_{N\to\infty} \lim_{M\to\infty} \sqrt{(1-\alpha)NM} \left(\widetilde{c}_\alpha^{N,M} - \widehat{c}_\alpha^N \right) \overset{\mathcal{D}}{\Rightarrow} \tau_{cvar} \mathcal{N}(0,1),
\end{cases} \tag{B.3.2}
$$

where

$$
\tau_{var} \overset{\triangle}{=} \sqrt{\mathbb{E}[\tau_\theta^2 | H(\theta) = v_\alpha]} \quad \text{and} \quad \tau_{cvar} \overset{\triangle}{=} \sqrt{\mathbb{E}[\tau_\theta^2 | H(\theta) \ge v_\alpha]}.
$$

Let us first establish (B.3.1). This is a direct result of Lemma B.1.1, where note that the order of A_N and B_N are strictly smaller than $O_{a.s.}(N^{-1/2})$. Furthermore, $H(\theta_i)$'s are i.i.d. random variables, and thus

$$\begin{aligned}\sigma^2_{var} &= NVar\left[\frac{1}{f(v_\alpha)}\left(\alpha - \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{H(\theta_i)\le v_\alpha\}\right)\right]\\ &= \frac{N}{f^2(v_\alpha)N}Var\left[\mathbb{1}\{H(\theta)\le v_\alpha\}\right] = \frac{\alpha(1-\alpha)}{f^2(v_\alpha)},\end{aligned}$$

and

$$\begin{aligned}\sigma^2_{cvar} &= NVar\left[\frac{1}{N}\sum_{i=1}^{N}\left[v_\alpha + \frac{1}{1-\alpha}\left(H(\theta_i)-v_\alpha\right)^+\right] - c_\alpha\right]\\ &= \frac{1}{(1-\alpha)^2}Var\left[\left(H(\theta)-v_\alpha\right)^+\right].\end{aligned}$$

Next, let us establish (B.3.2). For fixed N and scenarios $\theta_1, ..., \theta_N$, we have

$$\left(\widetilde{v}^{N,M}_\alpha - \widehat{v}^{N}_\alpha\right) = \widehat{H}_M(\theta^{(\alpha N)}) - H(\theta_{(\alpha N)}),$$

and

$$\left(\widetilde{c}^{N,M}_\alpha - \widehat{c}^{N}_\alpha\right) = \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}\widehat{H}_M(\theta^{(i)}) - \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N}H(\theta_{(i)}).$$

On the other hand, by Central Limit Theorem under Assumption 3.2.1.(ii),

$$\lim_{M\to\infty}\sqrt{M}\left(\widehat{H}_M(\theta_{(\alpha N)}) - H(\theta_{(\alpha N)})\right) \overset{\mathcal{D}}{\Rightarrow} \tau_{(\alpha N)}\mathcal{N}(0,1), \tag{B.3.3}$$

where $\tau^2_{(\alpha N)}$ is short for $\tau^2_{\theta_{(\alpha N)}} = Var[h(\theta_{(\alpha N)};\xi)|\theta_{(\alpha N)}]$. Note that

$$\sqrt{M}\left(\widetilde{v}^{N,M}_\alpha - \widehat{v}^{N}_\alpha\right) - \sqrt{M}\left(\widehat{H}_M(\theta_{(\alpha N)}) - H(\theta_{(\alpha N)})\right) = \sqrt{M}\left(\widehat{H}_M(\theta^{(\alpha N)}) - \widehat{H}_M(\theta_{(\alpha N)})\right).$$

By using the same technique in proving (B.1.9) and (B.1.10), we will show that

$$\lim_{M\to\infty}\sqrt{M}\left(\widehat{H}_M(\theta^{(\alpha N)}) - \widehat{H}_M(\theta_{(\alpha N)})\right) = 0. \quad w.p.1. \tag{B.3.4}$$

Indeed, denote the underlying sample space by Ω. In Lemma B.1.1, we have established that $\forall w \in \Omega$, there exists an M_ϵ such that $\forall M \ge M_\epsilon$, $(\theta^{(1)}_w, .., \theta^{(N)}_w) =$

$(\theta_{(1)}, ..., \theta_{(N)})$, where $\theta_w^{(i)}$ is the sample realization of $\theta^{(i)}$ with scenario w. It follows that $\forall M \geq M_\epsilon$,

$$\sqrt{M}\left(\widehat{H}_M^w(\theta_w^{(\alpha N)}) - \widehat{H}_M(\theta_{(\alpha N)})\right) = 0.$$

Thus, (B.3.4) holds. Combining with (B.3.3), we have

$$\sqrt{M}\left(\widehat{H}_M(\theta^{(\alpha N)}) - H(\theta_{(\alpha N)})\right) \overset{M\to\infty}{\Rightarrow} \tau_{(\alpha N)}\mathcal{N}(0,1). \tag{B.3.5}$$

Furthermore, similar to showing $\lim_{N\to\infty} \widehat{v}_\alpha^N = v_\alpha, w.p.1$, we can show that

$$\lim_{N\to\infty} \tau_{(\alpha N)}^2 = \sqrt{\mathbb{E}[\tau_\theta^2 | H(\theta) = v_\alpha]} = \tau_{var}^2.$$

Thus,

$$\lim_{N\to\infty}\lim_{M\to\infty} \sqrt{M}\left(\widehat{H}_M(\theta^{(\alpha N)}) - H(\theta_{(\alpha N)})\right) \overset{\mathcal{D}}{\Rightarrow} \tau_{var}\mathcal{N}(0,1),$$

and the first half of (B.3.2) holds. It remains to establish the second half of (B.3.2). Note that by Central Limit Theorem,

$$\lim_{M\to\infty} \sqrt{(1-\alpha)NM}\left(\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} \widehat{H}_M(\theta_{(i)}) - \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} H(\theta_{(i)})\right) \overset{\mathcal{D}}{\Rightarrow} \tau_I\mathcal{N}(0,1), \tag{B.3.6}$$

where $\tau_I \overset{\triangle}{=} \sqrt{\sum_{i=\alpha N}^{N} \tau_{(i)}^2 / [(1-\alpha)N]}$ and $\tau_{(i)}^2$ is short for $\tau_{\theta_{(i)}}^2$. Following a similar argument in showing (B.3.4), we can show

$$\lim_{M\to\infty} \sqrt{(1-\alpha)NM}\left(\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} \widehat{H}_M(\theta^{(i)}) - \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} \widehat{H}_M(\theta_{(i)})\right) = 0, \quad w.p.1. \tag{B.3.7}$$

Combining with (B.3.6), we have

$$\lim_{M\to\infty} \sqrt{(1-\alpha)NM}\left(\frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} \widehat{H}_M(\theta^{(i)}) - \frac{1}{(1-\alpha)N}\sum_{i=\alpha N}^{N} H(\theta_{(i)})\right) \overset{\mathcal{D}}{\Rightarrow} \tau_I\mathcal{N}(0,1). \tag{B.3.8}$$

Notice that

$$\lim_{N\to\infty} \tau_I^2 = \lim_{N\to\infty} \sum_{i=\alpha N}^{N} \frac{\tau_{(i)}^2}{(1-\alpha)N} = \mathbb{E}[\tau_\theta^2 | H(\theta) \geq v_\alpha] = \tau_{cvar}^2.$$

The latter half of (B.3.2) holds.

B.4 Proof of Theorem 3.2.4

Follow the notations in Appendix B.1 and B.2, we need to show that under Assumption 3.2.2, $N = o_M(M^2)$ is the sufficient and necessary condition for

$$\lim_{N,M\to\infty} \sqrt{N}\left(\widetilde{v}_\alpha^{N,M} - v_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{var}\mathcal{N}(0,1), \tag{B.4.1}$$

$$\lim_{N,M\to\infty} \sqrt{N}\left(\widetilde{c}_\alpha^{N,M} - c_\alpha\right) \overset{\mathcal{D}}{\Rightarrow} \sigma_{cvar}\mathcal{N}(0,1). \tag{B.4.2}$$

By Lemma B.2.2, B.2.3 and B.2.4, we have

$$\begin{aligned}\left(\widetilde{v}_\alpha^{N,M} - v_\alpha\right) &= \frac{1}{\widetilde{f}(\breve{v}_\alpha^M)}\left(\alpha - \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{\widehat{H}_M(\theta_i) \le \breve{v}_\alpha^M\}\right) \\ &\quad + \frac{-\Lambda'(v_\alpha)}{Mf(v_\alpha)} + o_M(\frac{1}{M}) + O_{a.s.}(N^{-3/4}(\log N)^{3/4}), \qquad \text{(B.4.3)} \\ \left(\widetilde{c}_\alpha^{N,M} - c_\alpha\right) &= \left(\frac{1}{N}\sum_{i=1}^{N}\left[\breve{v}_\alpha^M + \frac{1}{1-\alpha}\left(\widehat{H}_M(\theta_i) - \breve{v}_\alpha^M\right)^+\right] - \breve{c}_\alpha^M\right) \\ &\quad + \frac{\Lambda(v_\alpha)}{(1-\alpha)M} + o_M(\frac{1}{M}) + O_{a.s.}(N^{-1}\log N). \qquad \text{(B.4.4)}\end{aligned}$$

Note that $\breve{v}_\alpha^M \to v_\alpha$, $\breve{c}_\alpha^M \to c_\alpha$, and $\widetilde{f}(\breve{v}_\alpha^M) \to f(v_\alpha)$ as $M \to \infty$. We have

$$\lim_{M\to\infty} \frac{1}{\widetilde{f}(\breve{v}_\alpha^M)}\left(\alpha - \mathbb{1}\{\widehat{H}_M(\theta) \le \breve{v}_\alpha^M\}\right) = \frac{1}{f(v_\alpha)}\left(\alpha - \mathbb{1}\{H(\theta) \le v_\alpha\}\right), \quad w.p.1,$$

and

$$\begin{aligned}&\lim_{M\to\infty}\left(\breve{v}_\alpha^M + \frac{1}{1-\alpha}\left(\widehat{H}_M(\theta) - \breve{v}_\alpha^M\right)^+ - \breve{c}_\alpha^M\right) \\ =\ &\left(v_\alpha + \frac{1}{1-\alpha}\left(H(\theta) - v_\alpha\right)^+ - c_\alpha\right),\ w.p.1.\end{aligned}$$

Therefore, by Central Limit Theorem, (B.4.1) and (B.4.2) hold if and only if $N = o_M(M^2)$.

APPENDIX C

PROOFS OF THEOREMS IN CHAPTER IV

C.1 Proof of Theorems

Proof. **Proof of Lemma 4.3.1**. Since $S_{\theta_k}(\cdot)$ is continuous in both C_{α^*} and γ_{θ_k}, it suffices to show that for all $x \in \mathcal{X}$

$$\lim_{k\to\infty} \widehat{C}_{\alpha^*}(x) \to C_{\alpha^*}(x),\ w.p.1., \quad \text{and} \quad \lim_{k\to\infty} \widehat{\gamma}_{\theta_k} \to \gamma_{\theta_k},\ w.p.1. \tag{C.1.1}$$

Let us first show the left part of the above statement. Recall that by Assumption 4.3.1.(iii), we have $M_k \to \infty$ as $k \to \infty$. Then, we only need to show that the one-layer CVaR estimator $\widehat{C}_{\alpha^*}(x)$ is strongly consistent. By Lemma B.1.1 in Appendix B this holds, where note that Assumption 3.2.1 in Chapter 3.2 is satisfied by Assumption 4.3.2 here.

It remains to establish the right part of (C.1.1). In view of Assumption 4.3.1.(ii) and 4.3.1.(iii), we have $N_k, M_k \to \infty$ as $k \to \infty$. That is, N_k, M_k go to infinity simultaneously as $k \to$. Therefore, it suffices to show

$$\lim_{N_k, M_k\to\infty} \widehat{\gamma}_{\theta_k} \to \gamma_{\theta_k},\ w.p.1. \tag{C.1.2}$$

Note that

$$\gamma_{\theta_k} = VaR_{1-\rho}(-C_{\alpha^*}(x)),$$

i.e., the $(1-\rho)$-level Value-at-Risk (VaR) of $(-C_{\alpha^*}(x))$ w.r.t. $f(x;\theta_k)$. Furthermore,

$$\widehat{\gamma}_{\theta_k} = \widehat{VaR}_{1-\rho}(-\widehat{C}_{\alpha^*}(x)),$$

i.e., the sample $(1-\rho)$-quantile of $\{-\widehat{C}_{\alpha^*}(x_k^i) : i = 1, ..., N_k\}$. Therefore, $\widehat{\gamma}_{\theta_k}$ is a nested estimator of γ_{θ_k}, where N_k outer-layer samples are drawn, and for each outer-layer sample M_k inner-layer samples are drawn.

Rewrite $\widehat{C}_{\alpha^*}(x)$ as

$$\widehat{C}_{\alpha^*}(x) = C_{\alpha^*}(x) + \frac{1}{\sqrt{M_k}} \cdot \mathcal{E}_k(x), \quad \forall x \in \mathcal{X}, \tag{C.1.3}$$

where $\mathcal{E}_k(x)$ is the standardized error. By Theorem 3.2.3 in Chapter 3.2, we have

$$\lim_{M_k \to \infty} \sqrt{M_k} \left(\widehat{C}_{\alpha^*}(x) - C_{\alpha^*}(x) \right) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}\left(0, \sigma^2(x)\right),$$

where "$\overset{\mathcal{D}}{\Rightarrow}$" denotes the convergence in distribution, and $\mathcal{N}\left(0, \sigma^2(x)\right)$ denotes a normal distribution with mean zero and variance $\sigma^2(x)$, where $\sigma^2(x)$ is the variance parameter that only depends on x. Combined with (C.1.3), we can see that the standardized error $\mathcal{E}_k(x)$ converges to $\mathcal{N}\left(0, \sigma^2(x)\right)$ in distribution. Have establishing this, the remaining proof is identical to the proof of Theorem 3.2.2 in Chapter 3.2, where note that Assumption 4.3.2 here is parallel with Assumption 3.2.2 in Chapter 3.2. □

Proof. **Proof of Lemma 4.3.2**. With a slight abuse of notation, we also use $\|A\|_2$ to denote the spectral norm of a real square matrix A induced by the vector Euclidean norm. In particular, $\|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$, i.e., $\|A\|_2$ is the largest eigenvalue of the positive-semidefinite matrix $A^T A$. When the matrix A is positive-semidefinite, $\|A\|_2$ is just the largest eigenvalue of A.

To facilitate the proof, let us also introduce the following notations:

$$\widetilde{\mathbb{Y}}_k \triangleq \frac{1}{N_k} \sum_{i=1}^{N_k} S_{\theta_k}\left(-C_{\alpha^*}(x_k^i)\right) \Gamma(x_k^i), \;\; \widetilde{\mathbb{Z}}_k \triangleq \frac{1}{N_k} \sum_{i=1}^{N_k} S_{\theta_k}\left(-C_{\alpha^*}(x_k^i)\right),$$
$$\widehat{\mathbb{Y}}_k \triangleq \frac{1}{N_k} \sum_{i=1}^{N_k} \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^i)\right) \Gamma(x_k^i), \;\; \widehat{\mathbb{Z}}_k \triangleq \frac{1}{N_k} \sum_{i=1}^{N_k} \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^i)\right).$$

Here note that $\widetilde{\mathbb{Y}}_k$, $\widehat{\mathbb{Y}}_k$ are vectors because $\Gamma(\cdot)$ are vector-valued functions, and $\widetilde{\mathbb{Z}}_k$, $\widehat{\mathbb{Z}}_k$ are scalar-valued.

Since $C_{\alpha^*}(x)$ and $\Gamma(x)$ are both bounded on $\mathcal{X}$, we immediately have $|\widetilde{\mathbb{Z}}_k|$ bounded below from zero and $\frac{\left\|\widehat{\mathbb{Y}}_k\right\|_2}{|\widehat{\mathbb{Z}}_k|}$ bounded for all k. Note that

$$b_k \;\; = \;\; \widehat{V}_k^{-1} \left(\widehat{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] \right)$$

$$
\begin{aligned}
&= \widehat{V}_k^{-1}\left(\frac{\widehat{\mathbb{Y}}_k}{\widehat{\mathbb{Z}}_k} - \frac{\widetilde{\mathbb{Y}}_k}{\widetilde{\mathbb{Z}}_k}\right) \\
&= \widehat{V}_k^{-1}\left(\frac{\widehat{\mathbb{Y}}_k}{\widehat{\mathbb{Z}}_k} - \frac{\widehat{\mathbb{Y}}_k}{\widetilde{\mathbb{Z}}_k} + \frac{\widehat{\mathbb{Y}}_k}{\widetilde{\mathbb{Z}}_k} - \frac{\widetilde{\mathbb{Y}}_k}{\widetilde{\mathbb{Z}}_k}\right) \\
&= \widehat{V}_k^{-1}\widehat{\mathbb{Y}}_k\left(\frac{\widetilde{\mathbb{Z}}_k - \widehat{\mathbb{Z}}_k}{\widehat{\mathbb{Z}}_k\widetilde{\mathbb{Z}}_k}\right) + \widehat{V}_k^{-1}\frac{\widehat{\mathbb{Y}}_k - \widetilde{\mathbb{Y}}_k}{\widetilde{\mathbb{Z}}_k}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|b_k\|_2 &\leq \frac{\left\|\widehat{V}_k^{-1}\right\|_2}{\left|\widetilde{\mathbb{Z}}_k\right|}\frac{\left\|\widehat{\mathbb{Y}}_k\right\|_2}{\left|\widehat{\mathbb{Z}}_k\right|}\left|\widetilde{\mathbb{Z}}_k - \widehat{\mathbb{Z}}_k\right| + \frac{\left\|\widehat{V}_k^{-1}\right\|_2}{\left|\widetilde{\mathbb{Z}}_k\right|}\left|\widehat{\mathbb{Y}}_k - \widetilde{\mathbb{Y}}_k\right| \\
&\leq \frac{\left\|\widehat{V}_k^{-1}\right\|_2}{\left|\widetilde{\mathbb{Z}}_k\right|}\frac{\left\|\widehat{\mathbb{Y}}_k\right\|_2}{\left|\widehat{\mathbb{Z}}_k\right|}\frac{1}{N_k}\sum_{i=1}^{N_k}\left|S_{\theta_k}\left(-C_{\alpha^*}(x_k^i)\right) - \widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^i)\right)\right| \\
&\quad + \frac{\left\|\widehat{V}_k^{-1}\right\|_2}{\left|\widetilde{\mathbb{Z}}_k\right|}\frac{1}{N_k}\sum_{i=1}^{N_k}\left|\widehat{S}_{\theta_k}\left(-\widehat{C}_{\alpha^*}(x_k^i)\right) - S_{\theta_k}\left(-C_{\alpha^*}(x_k^i)\right)\right|\left\|\Gamma(x_k^i)\right\|_2.
\end{aligned}
$$

Recall that $\widehat{V}_k = \left(\widehat{Var}_{\theta_k}[\Gamma(x)] + \epsilon I\right)$. Thus, it is a positive-definite matrix and its minimum eigenvalue is at least ϵ. It follows that the maximum eigenvalue of $\widehat{V}_k^{-1}$ is no greater than ϵ^{-1}, i.e., $\left\|\widehat{V}_k^{-1}\right\|_2 \leq \epsilon^{-1}$. Since $|\widetilde{\mathbb{Z}}_k|$ is bounded below from zero, $\frac{\left\|\widehat{\mathbb{Y}}_k\right\|_2}{\left|\widehat{\mathbb{Z}}_k\right|}$ is bounded, and $\Gamma(x)$ is bounded on $\mathcal{X}$, Lemma 4.3.1 implies that $\|b_k\|_2 \to 0$ w.p.1 as $k \to \infty$. □

Proof. **Proof of Theorem 4.3.2**. Let us first show the following lemma.

Lemma C.1.1. *Suppose Assumption 4.3.1 and Assumption 4.3.2 hold. Further suppose the risk level sequence $\{\alpha_k\}$ generated by (4.2.1) converges to the target risk level α^* w.p.1. Then the sequence $\{\theta_k\}$ generated by (4.3.8) converges to a limit set of the ODE (4.3.7) w.p.1.*

Proof of Lemma C.1.1. Similar to the proof of Theorem 4.3.1, we will reformulate the updating scheme (4.3.8) as a noisy discretization of the constrained ODE (4.3.7), and show both the bias and the noise are properly bounded. Specifically, rewrite

(4.3.8) as

$$\theta_{k+1} = \theta_k + \beta_k \left[G(\theta_k) + \bar{b}_k + e_k + \bar{p}_k \right], \tag{C.1.4}$$

where $G(\theta_k)$ and e_k are defined as previously, $\bar{b}_k \triangleq \widehat{V}_k^{-1} \left(\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] \right)$, and $\bar{p}_k$ is the projection error term that takes the current iterate back onto the constraint set $\widetilde{\Theta}$ with minimum Euclidean norm. In view of Theorem 2 in [53], it suffices to show

$$\lim_{k\to\infty} \left\| \bar{b}_k \right\|_2 = 0, \quad w.p.1.$$

To ease the presentation, let us denote

$$\widetilde{\mathbb{E}}_{q_k}^{\alpha}[\Gamma(x)] \triangleq \sum_{i=1}^{N_k} \frac{S_{\theta_k}\left(-C_\alpha(x_k^i)\right)}{\sum_{j=1}^{N_k} S_{\theta_k}\left(-C_\alpha(x_k^j)\right)} \Gamma(x_k^i).$$

It immediately implies that $\widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] = \widetilde{\mathbb{E}}_{q_k}^{\alpha^*}[\Gamma(x)]$. Furthermore,

$$\begin{aligned}
\left\| \bar{b}_k \right\|_2 &= \left\| \widehat{V}_k^{-1} \left(\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}[\Gamma(x)] \right) \right\|_2 \\
&= \left\| \widehat{V}_k^{-1} \left(\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\alpha_k}[\Gamma(x)] \right) + \widehat{V}_k^{-1} \left(\widetilde{\mathbb{E}}_{q_k}^{\alpha_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\alpha^*}[\Gamma(x)] \right) \right\|_2 \\
&\leq \left\| \widehat{V}_k^{-1} \left(\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\alpha_k}[\Gamma(x)] \right) \right\|_2 + \left\| \widehat{V}_k^{-1} \right\|_2 \left\| \widetilde{\mathbb{E}}_{q_k}^{\alpha_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\alpha^*}[\Gamma(x)] \right\|_2 .
\end{aligned} \tag{C.1.5}$$

Following an argument almost identical to the proof of Lemma 4.3.2, the first term in (C.1.5) converges to 0 w.p.1 as $k \to \infty$. Note that $S_{\theta_k}(\cdot)$ is a continuous function and $C_{\alpha_k}(x)$ is continuous in α_k. Thus, $\widetilde{\mathbb{E}}_{q_k}^{\alpha_k}[\Gamma(x)]$ is a continuous function in α_k. Therefore, the second term in (C.1.5) converges to 0 w.p.1 as $k \to \infty$ since $\left\| \widehat{V}_k^{-1} \right\|_2$ is bounded and α_k converges to α^* as $k \to \infty$. Proof of Lemma C.1.1 is now complete.

In view of Lemma C.1.1, it remains to show that the risk level sequence $\{\alpha_k\}$ generated by (4.2.1) converges to the target risk level α^* w.p.1. Proof by contradiction. Since the sequence $\{\alpha_k\}$ is non-decreasing and bounded above by α^*, let us assume $\lim_{k\to\infty} \alpha_k = \bar{\alpha}^*$ and $\bar{\alpha}^* < \alpha^*$ w.p.1. Conditioning on this, Lemma C.1.1 still holds when the target risk level α^* is replaced by $\bar{\alpha}^*$. That is, the algorithm GASS-CVaR-ARL converges, and the gradient sequence $\{g_k\}$ converges to 0 w.p.1. as $k \to \infty$. Note that g_k is bounded (since $\Gamma(x)$ is bounded), by bounded convergence theorem

we have

$$\lim_{k\to\infty} \mathbb{E}\left[\|g_k\|_2\right] = 0. \tag{C.1.6}$$

Furthermore, note that

$$\begin{aligned}\mathbb{E}\left[\|\bar{g}_k - g_k\|_2\right] =& \mathbb{E}\left[\left\|\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right] \\ \leq& \mathbb{E}\left[\left\|\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right] + \mathbb{E}\left[\left\|\widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)] - \mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right],\end{aligned} \tag{C.1.7}$$

where

$$\mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)] \triangleq \frac{\int S_{\theta_k}\left(-C_{\bar{\alpha}^*}(x)\right)\Gamma(x) f(x;\theta_k)dx}{\int S_{\theta_k}\left(-C_{\bar{\alpha}^*}(x)\right) f(x;\theta_k)dx}.$$

We have shown in the proof of Lemma C.1.1 that

$$\lim_{k\to\infty}\left\|\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2 = 0, \quad w.p.1.$$

Since $\left\|\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2$ is bounded, again by bounded convergence theorem

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|\overline{\mathbb{E}}_{q_k}[\Gamma(x)] - \widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right] = 0. \tag{C.1.8}$$

Moreover, notice that $\widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]$ is a self-normalized importance sampling estimator of $\mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]$. Applying Theorem 9.1.10 (pp. 294) in [25], we have

$$\mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma_j(x)] - \mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma_j(x)]\right|^2\right] \leq \frac{c_j}{N_k},\ j = 1, ..., d_\theta,$$

where $\Gamma_j(x)$ is the j^{th} element in the vector $\Gamma(x)$, and c_j's are positive constants that depend on the bounds of $\Gamma_j(x)$'s on $\mathcal{X}$. Therefore, by Cauchy-Schwarz Inequality we have

$$\mathbb{E}\left[\left\|\widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)] - \mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right] \leq \sqrt{\frac{d\cdot\max_j c_j}{N_k}}.$$

That is,

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|\widetilde{\mathbb{E}}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)] - \mathbb{E}_{q_k}^{\bar{\alpha}^*}[\Gamma(x)]\right\|_2\right] = 0. \tag{C.1.9}$$

Combining (C.1.7), (C.1.8) with (C.1.9), we have

$$\lim_{k\to\infty} \mathbb{E}\left[\|\bar{g}_k - g_k\|_2\right] = 0.$$

In view of (C.1.6), we have

$$\lim_{k\to\infty} \mathbb{E}\left[\|\bar{g}_k\|_2\right] = 0. \tag{C.1.10}$$

Since $\overline{\alpha}^* < \alpha^*$, the sequence $\{\|\bar{g}_k\|_2\}$ generated by (4.2.1) will always be above a certain positive value w.p.1 (otherwise α_k will converge to α^*), which contradicts with (C.1.10). Proof is complete. □

REFERENCES

[1] Alagoz, O., Hsu, H., Schaefer, A. J., and Roberts, M. S., "Markov decision processes: a tool for sequential decision making under uncertainty," *Medical Decision Making*, 2009.

[2] Alexander, S., Coleman, T. F., and Li, Y., "Minimizing cvar and var for a portfolio of derivatives," *Journal of Banking & Finance*, vol. 30, no. 2, pp. 583–605, 2006.

[3] Andersen, L. and Broadie, M., "Primal-dual simulation algorithm for pricing multidimensional American options," *Management Science*, vol. 50, no. 9, pp. 1222 – 1234, 2004.

[4] Ankenman, B., Nelson, B. L., and Staum, J., "Stochastic kriging for simulation metamodeling," *Operations Research*, vol. 58, no. 2, pp. 371–382, 2010.

[5] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D., "Coherent measures of risk," *Mathematical Finance*, vol. 9, pp. 203–228, 1999.

[6] Barton, R. R., "Tutorial: Input uncertainty in outout analysis," in *Proceedings of the 2012 Winter Conference on Simulation*, pp. 1–12, 2012.

[7] Barton, R. R., Nelson, B. L., and Xie, W., "Quantifying input uncertainty via simulation confidence intervals," *INFORMS Journal on Computing*, vol. 26, no. 1, pp. 74–87, 2013.

[8] Barton, R. R. and Schruben, L. W., "Uniform and bootstrap resampling of empirical distributions," in *Proceedings of the 1993 Winter Conference on Simulation*, pp. 503–508, 1993.

[9] Barton, R. R. and Schruben, L. W., "Resampling methods for input modeling," in *Proceedings of the 2001 Winter Conference on Simulation*, pp. 372–378, 2001.

[10] Bäuerle, N. and Rieder, U., *Markov Decision Processes with Applications to Finance*. Springer Science & Business Media, 2011.

[11] Bellman, R., *Dynamic Programming*. Dover Publications, 1957.

[12] Belomestny, D., Bender, C., and Schoenmakers, J., "True upper bounds for Bermudan products via non-nested Monte Carlo," *Mathematical Finance*, vol. 19, pp. 53 – 71, 2009.

[13] Bertsekas, D. P. and Tsitsiklis, J. N., *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 2nd ed., 1997.

[14] Bertsekas, D., *Dynamic Programming and Optimal Control*. Athena Scientific, 3rd ed., 2007.

[15] Biller, B. and Corlu, C. G., "Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs," *Operations Research*, vol. 59, no. 3, pp. 661–673, 2011.

[16] Bogachev, V. I., *Measure Theory*, vol. 1. Springer Science & Business Media, 2007.

[17] Borkar, V. S., *Stochastic Approximation: A Dynamical Systems ViewPoint*. Cambridge University Press, 2008.

[18] Broadie, M., Du, Y., and Moallemi, C. C., "Efficient risk estimation via nested sequential simulation," *Management Science*, vol. 57, no. 6, pp. 1172–1194, 2011.

[19] Broadie, M., Du, Y., and Moallemi, C. C., "Risk estimation via regression," *Working Paper*, 2015.

[20] Brown, D. B. and Haugh, M. B., "Information relaxation bounds for infinite horizon markov decision processes," *Working paper*, 2016.

[21] Brown, D. B. and Smith, J. E., "Optimal sequential exploration: Bandits, clairvoyants, and wildcats," *Operations Research*, vol. 61, no. 3, pp. 644–665, 2013.

[22] Brown, D. B. and Smith, J. E., "Information relaxations, duality, and convex stochastic dynamic programs," *Operations Research*, vol. 62, no. 6, pp. 1394–1415, 2014.

[23] Brown, D. and Smith, J., "Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds," *Management Science*, vol. 57, no. 10, pp. 1752–1770, 2011.

[24] Brown, D., Smith, J., and Sun, P., "Information relaxations and duality in stochastic dynamic programs," *Operations Research*, vol. 58, no. 4, pp. 758 – 801, 2010.

[25] Cappé, O., Moulines, E., and Rydén, T., *Inference in Hidden Markov Models*. Springer Series in Statistics, New York: Springer, 2005.

[26] Chang, H. S., Hu, J., Fu, M. C., and Marcus, S. I., *Simulation-based Algorithms for Markov Decision Processes*. Communications and Control Engineering Series, Springer, 2013.

[27] Chen, N. and Glasserman, P., "Additive and multiplicative duals for american option pricing," *Finance and Stochastics*, vol. 11, pp. 153 – 179, 2007.

[28] Cheng, R. C. and Holland, W., "Sensitivity of computer simulation experiments to errors in input data," *Journal of Statistical Computation and Simulation*, vol. 57, no. 1-4, pp. 219–241, 1997.

[29] Chick, S. E., "Input distribution selection for simulation experiments: accounting for input uncertainty," *Operations Research*, vol. 49, no. 5, pp. 744–758, 2001.

[30] Collin-Dufresne, P., Daniel, K., Moallemi, C., and Saglam, M., "Strategic asset allocation with predictable returns and transaction costs," *Working Paper*, 2015.

[31] Davis, M. and Burstein, G., "A deterministic approach to stochastic optimal control with application to anticipative control," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 40, no. 3-4, pp. 203–256, 1992.

[32] Davis, M. and Zervos, M., "A new proof of the discrete-time LQG optimal control theorems," *IEEE Transactions on Autonmatic Control*, vol. 40, no. 8, pp. 1450–1453, 1995.

[33] de Farias, D. and van Roy, B., "The linear programming approach to approximate dynamic programming," *Operations Research*, pp. 850–865, 2003.

[34] Degeilh, Y. and Gross, G., "Stochastic simulation of utility-scale storage resources in power systems with integrated renewable resources," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1424–1434, 2015.

[35] Desai, V., Farias, V., and Moallemi, C., "Bounds for Markov decision processes." Chapter in Reinforcement Learning and Approximate Dynamic Programming for Feedback Control (F. L. Lewis, D. Liu, eds.), 2011.

[36] Desai, V. V., Farias, V. F., and Moallemi, C. C., "Pathwise optimization for optimal stopping problems," *Management Science*, vol. 58, no. 12, pp. 2292–2308, 2012.

[37] Dorigo, M. and Blum, C., "Ant colony optimization theory: A survey," *Theoretical Computer Science*, vol. 344, no. 2, pp. 243–278, 2005.

[38] Gârleanu, N. and Pedersen, L. H., "Dynamic trading with predictable returns and transaction costs," *The Journal of Finance*, vol. 68, no. 6, pp. 2309–2340, 2013.

[39] Glasserman, P., Heidelberger, P., and Shahabuddin, P., "Variance reduction techniques for estimating value-at-risk," *Management Science*, vol. 46, no. 10, pp. 1349–1364, 2000.

[40] Gordy, M. B. and Juneja, S., "Nested simulation in portfolio risk measurement," *Management Science*, vol. 56, no. 10, pp. 1833–1848, 2010.

[41] Haugh, M. and Wang, C., "Dynamic portfolio execution and information relaxations," *SIAM Journal on Financial Mathematics*, vol. 5, no. 1, pp. 316–359, 2014.

[42] Haugh, M. and Wang, C., "Information relaxations and dynamic zero-sum games," *Working Paper*, 2014.

[43] Haugh, M. B. and Jain, A., "The dual approach to portfolio evaluation: A comparison of the static, myopic and generalized buy-and-hold strategies," *Quantitative Finance*, vol. 11, no. 1, pp. 81–99, 2011.

[44] Haugh, M. and Kogan, L., "Pricing American options: A duality approach," *Operations Research*, vol. 52, no. 2, pp. 258 – 270, 2004.

[45] Haugh, M. and Lim, A., "Linear-quadratic control and information relaxations," *Operations Research Letters*, vol. 40, no. 6, pp. 521 – 528, 2012.

[46] Henderson, S. G., "Input model uncertainty: Why do we care and what should we do about it?," in *Proceedings of the 2003 Winter Conference on Simulation*, pp. 90–100, 2003.

[47] Hong, L. J., "Estimating quantile sensitivities," *Operations Research*, vol. 57, no. 1, pp. 118–130, 2009.

[48] Hong, L. J., Hu, Z., and Liu, G., "Monte carlo methods for value-at-risk and conditional value-at-risk: A review," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 24, no. 4, pp. 1–37, 2014.

[49] Hong, L. J. and Liu, G., "Simulating sensitivities of conditional value at risk," *Management Science*, vol. 55, no. 2, pp. 281–293, 2009.

[50] Hu, J., Fu, M. C., and Marcus, S. I., "A model reference adaptive search method for global optimization," *Operations Research*, vol. 55, no. 3, pp. 549–568, 2007.

[51] Hu, J., Fu, M. C., Marcus, S. I., and others, "A model reference adaptive search method for stochastic global optimization," *Communications in Information & Systems*, vol. 8, no. 3, pp. 245–276, 2008.

[52] Kim, M. J. and Lim, A. E., "Robust multiarmed bandit problems," *Management Science*, vol. 62, no. 1, pp. 264–285, 2015.

[53] Kushner, H., "Stochastic approximation: a survey," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 87–96, 2010.

[54] Kushner, H. and Yin, G. G., *Stochastic Approximation and Recursive Algorithms and Applications*, vol. 35. Springer Science & Business Media, 2003.

[55] Kushner, H. J. and Clark, D. S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, vol. 26. Springer Science & Business Media, 2012.

[56] Lai, G., Margot, F., and Secomandi, N., "An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation," *Operations Research*, vol. 58, no. 3, pp. 564–582, 2010.

[57] Lai, G., Wang, M. X., Kekre, S., Scheller-Wolf, A., and Secomandi, N., "Valuation of storage at a liquefied natural gas terminal," *Operations Research*, vol. 59, no. 3, pp. 602–616, 2011.

[58] Lan, H., Nelson, B. L., and Staum, J., "A confidence interval procedure for expected shortfall risk measurement via two-level simulation," *Operations Research*, vol. 58, no. 5, pp. 1481–1490, 2010.

[59] Larranaga, P. and Lozano, J. A., *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*, vol. 2. Springer Science & Business Media, 2002.

[60] Lee, S.-H., *Monte Carlo computation of conditional expectation quantiles*. PhD thesis, Stanford University, Stanford, USA., 1998.

[61] Liu, M. and Staum, J., "Stochastic kriging for efficient nested simulation of expected shortfall," *Journal of Risk*, vol. 12, no. 3, pp. 3–27, 2010.

[62] Moallemi, C. C. and Saglam, M., "Dynamic portfolio choice with linear rebalancing rules," *Working Paper*, 2015.

[63] Molvalioglu, O., Zabinsky, Z. B., and Kohn, W., "The interacting-particle algorithm with dynamic heating and cooling," *Journal of Global Optimization*, vol. 43, no. 2-3, pp. 329–356, 2009.

[64] Molvalioglu, O., Zabinsky, Z. B., and Kohn, W., "Meta-control of an interacting-particle algorithm for global optimization," *Nonlinear Analysis: Hybrid Systems*, vol. 4, no. 4, pp. 659–671, 2010.

[65] Powell, W., *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley and Sons, 2nd ed., 2011.

[66] Puterman, M. L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley & Sons, 1994.

[67] Rockafellar, R. T. and Uryasev, S., "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–42, 2000.

[68] Rockafellar, R. T. and Uryasev, S., "Conditional value-at-risk for general loss distributions," *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[69] Rogers, L. C. G., "Pathwise stochastic optimal control," *SIAM Journal on Control and Optimization*.

[70] Rogers, L. C. G., "Monte Carlo valuation of American options," *Mathematical Finance*, vol. 12, no. 3, pp. 271 – 286, 2002.

[71] Romeijn, H. E. and Smith, R. L., "Simulated annealing for constrained global optimization," *Journal of Global Optimization*, vol. 5, no. 2, pp. 101–126, 1994.

[72] Rouvinez, C., "Going greek with var," *Risk*, vol. 10, no. 2, pp. 57–63, 1997.

[73] Rubinstein, R. Y., "Combinatorial optimization, cross-entropy, ants and rare events," *Stochastic Optimization: Algorithms and Applications*, vol. 54, pp. 303–363, 2001.

[74] Ruszczyński, A., "Risk-averse dynamic programming for markov decision processes," *Mathematical Programming*, vol. 125, no. 2, pp. 235–261, 2010.

[75] Ruszczyński, A. and Shapiro, A., "Optimization of convex risk functions," *Mathematics of Operations Research*, vol. 31, no. 3, pp. 433–452, 2006.

[76] Sarimveis, H., Patrinos, P., Tarantilis, C. D., and Kiranoudis, C. T., "Dynamic modeling and control of supply chain systems: A review," *Computers & Operations Research*, vol. 35, no. 11, pp. 3530–3561, 2008.

[77] Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, vol. 162. John Wiley & Sons, 2009.

[78] Shapiro, A., Dentcheva, D., and others, *Lectures on stochastic programming: modeling and theory*, vol. 16. SIAM, 2014.

[79] Steckley, G., *Estimating the density of a conditional expectation*. PhD thesis, Cornell University, Ithaca, NY, 2006.

[80] Sun, L. and Hong, L. J., "Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk," *Operations Research Letters*, vol. 38, no. 4, pp. 246–251, 2010.

[81] Sun, Y., Apley, D. W., and Staum, J., "Efficient nested simulation for estimating the variance of a conditional expectation," *Operations Research*, vol. 59, no. 4, pp. 998–1007, 2011.

[82] Xie, W., Nelson, B. L., and Barton, R. R., "A bayesian framework for quantifying uncertainty in stochastic simulation," *Operations Research*, vol. 62, no. 6, pp. 1439–1452, 2014.

[83] Xie, W., Nelson, B. L., and Barton, R. R., "Statistical uncertainty analysis for stochastic simulation," *Working Paper*, 2015.

[84] Ye, F. and Zhou, E., "Parameterized penalties in the dual representation of Markov decision processes," in *Proceedings of the 51st IEEE Conference on Decision and Control*, pp. 870–876, 2012.

[85] Ye, F., Zhu, H., and Zhou, E., "Weakly coupled dynamic program: Information and lagrangian relaxations," *Working Paper*, 2016.

[86] Ye, F. and Zhou, E., "Information relaxation and dual formulation of controlled markov diffusions," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2676–2691, 2015.

[87] Zhou, E. and Hu, J., "Gradient-based adaptive stochastic search for non-differentiable optimization," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1818–1832, 2014.

[88] Zhu, H., Hale, J., and Zhou, E., "Risk optimization via gradient-based adaptive stochastic search," *Working Paper*, 2016.

[89] Zhu, H., Ye, F., and Zhou, E., "Fast estimation of true bounds on bermudan option prices under jump-diffusion processes," *Quantitative Finance*, vol. 15, no. 11, pp. 1885–1900, 2015.

[90] Zhu, H., Ye, F., and Zhou, E., "Solving the dual problems of dynamic programs via regression," *Working Paper*, 2016.

[91] Zhu, H. and Zhou, E., "Risk assessment of input uncertainty," *Working Paper*, 2016.

[92] Zouaoui, F. and Wilson, J. R., "Accounting for parameter uncertainty in simulation input modeling," *IIE Transactions*, vol. 35, no. 9, pp. 781–792, 2003.

[93] Zouaoui, F. and Wilson, J. R., "Accounting for input-model and input-parameter uncertainties in simulation," *IIE Transactions*, vol. 36, no. 11, pp. 1135–1151, 2004.